

CC obid. by mutagenesis of the C-terminal region (AAR75500), exhibit
 CC excellent antagonistic properties and substantially no agonist
 CC activity. Preferred analogs are C5a(1-72,L72C), C5a(1-71,Q71C),
 CC C5a(1-70,M70C), C5a(1-69,D69C) and C5a(1-68,K68C).
 XX
 SO Sequence 11 AA:

Query Match 100.0%; Score 52; DB 16; Length 11;
 Best Local Similarity 100.0%; Pred. No. 0.00018;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Caps 0;
 Oy 1 ISHKDMQLGR 10
 Db 2 ISHKDMQLGR 11

RESULT 2
 AAB74056
 ID AAB74056 standard; Peptide: 20 AA.
 XX
 AC AAB74056:
 XX
 DT 16-MAY-2001 (first entry)
 XX
 DE Human C5a peptide fragment #3.
 XX
 KW Human: C5a; complement; antibody; bacterial infection; sinusitis;
 KW meningitis; respiratory; gastrointestinal; urinary tract infection;
 KW wound; anaphylatoxin; sepsis.
 XX
 OS Homo sapiens.
 XX
 PN WO200115731-A1.
 PD 08-MAR-2001.
 XX
 PF 31-AUG-2000: 2000MO-US24219.
 XX
 PR 31-AUG-1999: 99US-0387671.
 XX
 PA (UNMI) UNIV MICHIGAN.
 XX
 PI Ward PA, Huber-Lang M, Sarma V:
 XX
 DR WPI: 2001-226665/23.
 XX

PT Compositions for treating blood-borne and toxin mediated diseases and
 PT treatment of sepsis in humans and other animals comprises anti-C5a
 PT antibodies generated against C-terminal truncated C5a peptides -
 XX
 PS Example 7; Page 26; 84pp; English.
 XX

CC The present sequence is a peptide fragment of human complement component
 CC C5a (the full-length sequence is given in AAB74053). The present
 CC invention relates to an antibody specific for the present sequence. The
 CC C5a-antibody can be used in a therapeutic composition, which is useful
 CC for treating a subject suffering from bacterial infection, e.g.
 CC sinusitis, meningitis, respiratory, gastrointestinal or urinary tract
 CC infections or infections in wounds. In addition, the C5a antibody can
 CC be used for treating sepsis. C5a is also known as anaphylatoxin.
 CC
 XX
 SO Sequence 20 AA:

Query Match 100.0%; Score 52; DB 22; Length 20;
 Best Local Similarity 100.0%; Pred. No. 0.00036;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Caps 0;
 Oy 1 ISHKDMQLGR 10
 Db 11 ISHKDMQLGR 20

RESULT 3
 AAP71666
 ID AAP71666 standard; Protein: 74 AA.
 XX
 AC AAP71666:
 XX
 DT 25-APR-1991 (first entry)
 XX
 DE Human anaphylatoxin C5a.
 XX
 KW activated complement component C5a.
 XX
 OS Synthetic.
 XX
 PN EP220864-A.
 XX
 PD 06-MAY-1987.
 XX
 PF 10-OCT-1986: 86EP-0307839.
 XX
 PR 18-OCT-1985: 85US-0789206.
 PR 18-MAR-1986: 86US-0841121.
 XX
 PA (PEIZ) PEIZER INC.
 XX
 PI Davidow LS, Franke AE, Dezeuw JR:
 XX
 DR WPI: 1987-124409/18.
 XX

PT New Yarrowia lipolytica transformants - used for expression and
 PT secretion of heterologous proteins, esp. protenimin, and human
 PT anaphylatoxin C5a
 XX

PS Example: Fig 9; 45pp; English.
 XX

CC The synthetic gene encoding this protein was constructed from ten
 CC 47-mer oligonucleotides. The sequence was chosen to maximise E. coli
 CC and S.cerevisiae preferred codon utilisation and to allow for
 CC several restriction enzyme sites. Direct expression of C5a in
 CC E.coli is possible by inclusion of an ATG codon in front of the
 CC triplec coding for the first amino acid of the C5a polypeptide.
 CC See also AAN70213-6, AAN70218, AAN71339, AAN71340, AAN71343-8.
 XX
 SO Sequence 74 AA:

Query Match 100.0%; Score 52; DB 8; Length 74;
 Best Local Similarity 100.0%; Pred. No. 0.0016;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Caps 0;
 Oy 1 ISHKDMQLGR 10
 Db 65 ISHKDMQLGR 74

RESULT 4
 AAR75497
 ID AAR75497 standard; Protein: 74 AA.
 XX
 AC AAR75497:
 XX
 DT 13-NOV-1995 (first entry)
 XX
 DE Human C5a protein.
 XX
 KW C5a; C5a receptor-antagonist; antiinflammatory.
 XX
 OS Synthetic.
 XX
 PN WO9516033-A.
 XX
 PD 15-JUN-1995.
 XX

XX 16-NOV-1994: 94MO-IB00359.
 XX 06-DEC-1993: 93US-0162591.
 XX (CIBA) CIBA GEIGY AG.
 XX Boyar WC, Galakatos NG, Peppard JV, Van Oostrum J;
 PI WPI: 1995-224319/29.
 DR N-PSDB: AA092518.
 XX C5a receptor antagonists having no agonist activity - are used in
 PT compns. to treat C5a-mediated diseases and inflammatory conditions
 XX
 PS Disclosure: Page 36-37: 65pp: English.
 XX C5a encoded by a synthetic gene is given in AAR75497. Analogs of C5a,
 CC obt'd. by mutagenesis of the C-terminal region, exhibit excellent
 CC antagonistic properties and substantially no agonist activity.
 CC
 SO Sequence 74 AA:

Query Match 100.0%; Score 52; DB 16; Length 74;
 Best Local Similarity 100.0%; Pred. No. 0.0016;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ISHKDMQLGR 10
 |||||
 DB 65 Ishkdmqlgr 74

RESULT 5
 AAW99580
 ID AAW99580 standard; peptide: 74 AA.
 AC AAW99580;
 XX 22-JUN-1999 (first entry)
 DE Solid phase sequential ligation of complement factor 5A.
 XX Sequential ligation; immobilisation: thioacid; thioester: library.
 KW Synthetic.
 OS
 XX Key Location/Qualifiers
 FT Region 15 /note= "designated J in specification"
 FT
 XX WO9856807-A1.
 PN 17-DEC-1998.
 PD 12-JUN-1998: 98WO-US12278.
 XX 13-JUN-1997: 97US-0049553.
 PR (GRYP-) GRYPHON SCI.
 PA Canne L, Kent SBH, Simon R;
 XX WPI: 1999-142422/12.
 DR
 XX Solid phase assembly of long peptides using thioester to react with
 PT terminal cysteine - in unprotected peptide segments, provide rapid
 PT and selective reaction without need for intermediate purification,
 PT permits synthesis in N to C or C to N directions
 XX Example 4: Page 31: 87pp: English.
 PS This sequence represents the complement factor 5a protein (amino acids
 CC

CC 1-74) generated by sequential ligation of 3 peptides comprising amino
 CC acids 1-20, 21-46 and 47-74 by the method of the invention. This involves
 CC immobilising a peptide on a solid surface via a cleavable linker and
 CC sequentially ligating peptides containing a N-terminal Cys residue and
 CC a thioacid at the C-terminus. Once ligated to the immobilised peptide,
 CC the thioacid is converted to a thioester, ready for ligation of the
 CC next C-terminal Cys-containing peptide. Optionally the first peptide
 CC (immobilised on the solid support) contains a thioacid at the C-terminus
 CC which is converted to a thioester prior to ligation of the second
 CC peptide. The method is specifically used to produce polypeptide
 CC libraries.
 CC
 SO Sequence 74 AA:

Query Match 100.0%; Score 52; DB 20; Length 74;
 Best Local Similarity 100.0%; Pred. No. 0.0016;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ISHKDMQLGR 10
 |||||
 DB 65 Ishkdmqlgr 74

RESULT 6
 AAE05454
 ID AAE05454 standard; Protein: 74 AA.
 AC AAE05454;
 XX 24-SEP-2001 (first entry)
 DE Human C5a anaphylatoxin.
 XX Mannan binding protein-associated serine protease: MASP.
 KW Complement-activation; C-activation; microbial infection;
 KW cytomagalovirus; CMV; hepatitis virus; human immunodeficiency virus;
 KW HIV; organ transplant rejection; tissue injury; autoimmune disease;
 KW rheumatoid arthritis; systemic lupus erythematosus; SLE; human;
 KW inflammatory response; Alzheimer's disease; C5a anaphylatoxin.
 XX
 OS Homo sapiens.
 XX
 FT Key Location/Qualifiers
 FT Modified-site 64..66 /note= "Asn is N-glycosylated"
 FT
 XX US6235494-B1.
 PN 22-MAY-2001.
 PD 08-FEB-1999: 99US-0246500.
 XX 08-FEB-1999: 99US-0246500.
 PR (SCRI) SCRIPPS RES INST.
 PA Hugli TE;
 XX WPI: 2001-450082/48.
 DR
 XX Method for determining in-vivo levels of activated mannan binding
 PT protein-associated serine protease enzyme, and new peptide derivatives
 PT
 XX Claim 3: Column 47-50: 39pp: English.
 PS The invention relates to assays for measuring in-vivo levels of
 CC activated mannan-binding protein-associated serine protease (MASP-1) and
 CC MASP-2) activity. The assay comprises contacting a sample of blood or
 CC plasma comprising a metal ion chelator with a substrate of formula:
 CC R-peptide-Y; wherein the peptide comprises at least 4 residues from the
 CC C-terminus of an anaphylatoxin selected from C3a, C4a and C5a; R is a

CC peptide blocking, protecting or capping group; and Y is any group
 CC cleavable from the substrate by MASP and comprises a labeled tag. The
 CC substrate of the invention is also useful for monitoring in vitro and
 CC in vivo complement-activation (C-activation) by classical, alternative
 CC or lectin pathways. The assays of MASP activity can be used for
 CC detecting or monitoring a condition associated with complement
 CC activation. The conditions include microbial infections
 CC caused by particularly cytomegalovirus (CMV), hepatitis virus and
 CC human immunodeficiency virus (HIV), organ transplant rejection, tissue
 CC injury, autoimmune diseases (e.g. rheumatoid arthritis, systemic lupus
 CC erythematosus (SLE)) and inflammatory responses (e.g. in Alzheimer's
 CC disease and bacterial diseases). The assay of MASP is also useful for
 CC assessing the toxicity or injury of therapeutic treatments or screening
 CC test compounds as agents for treatment of viral diseases, parasitic
 CC infections, tissue injury, organ transplant rejection, autoimmune
 CC diseases or inflammatory responses. The present sequence is human
 CC C5a anaphylatoxin.

CC Sequence 74 AA:

Query Match 100.0%; Score 52; DB 22; Length 74;

Best Local Similarity 100.0%; Pred. No. 0.0016;

Matches 10: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ISHKDMOLGR 10
 |||||
 Db 65 lshkdmqlgr 74

RESULT 7

AAB74053 AAB74053 standard; protein: 74 AA.

AC AAB74053:

DT 16-MAY-2001 (first entry)

DE Human C5a.

XX Human; C5a: complement; antibody; bacterial infection; sinusitis;

KW meningitis; respiratory; gastrointestinal; urinary tract infection;

KW wound; anaphylatoxin; sepsis.

OS Homo sapiens.

XX WO200115731-A1.

PD 08-MAR-2001.

PF 31-AUG-2000; 2000WO-US24219.

PR 31-AUG-1999; 99US-0387671.

PA (UNMT) UNIV MICHIGAN.

PI Ward PA, Huber-Lang M, Sarma V;

DR WPI: 2001-226665/23.

XX N-PSDB: AAF75791.

PT Compositions for treating blood-borne and toxin mediated diseases and

PT treatment of sepsis in humans and other animals comprises anti-C5a

XX antibodies generated against C-terminal truncated C5a peptides

PS Example 7; Page 26; 84pp; English.

XX The present sequence is human complement component C5a. The present

CC invention relates to an antibody specific for the present sequence. The

CC C5a-antibody can be used in a therapeutic composition, which is useful

CC for treating a subject suffering from bacterial infection, e.g.

CC sinusitis, meningitis, respiratory, gastrointestinal or urinary tract

CC infections or infections in wounds. In addition, the C5a antibody can

CC be used for treating sepsis. C5a is also known as anaphylatoxin.

XX Sequence 74 AA:

Query Match 100.0%; Score 52; DB 22; Length 74;

Best Local Similarity 100.0%; Pred. No. 0.0016;

Matches 10: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ISHKDMOLGR 10
 |||||
 Db 65 lshkdmqlgr 74

RESULT 8

AAB74119 AAB74119 standard; protein: 74 AA.

AC AAB74119:

DT 16-MAY-2001 (first entry)

DE Variant human C5a.

XX Human; C5a: complement; antibody; bacterial infection; sinusitis;

KW meningitis; respiratory; gastrointestinal; urinary tract infection;

KW wound; anaphylatoxin; sepsis.

OS Homo sapiens.

XX Key Location/Qualifiers

FT MISC-difference 27 /Label- Cys, Ser

PN WO200115731-A1.

PD 08-MAR-2001.

PF 31-AUG-2000; 2000WO-US24219.

PR 31-AUG-1999; 99US-0387671.

PA (UNMT) UNIV MICHIGAN.

PI Ward PA, Huber-Lang M, Sarma V;

DR WPI: 2001-226665/23.

PT Compositions for treating blood-borne and toxin mediated diseases and

PT treatment of sepsis in humans and other animals comprises anti-C5a

XX antibodies generated against C-terminal truncated C5a peptides

PS Disclosure: Fig 8; 84pp; English.

XX The present sequence is human complement component C5a. The present

CC invention relates to an antibody specific for the present sequence. The

CC C5a-antibody can be used in a therapeutic composition, which is useful

CC for treating a subject suffering from bacterial infection, e.g.

CC sinusitis, meningitis, respiratory, gastrointestinal or urinary tract

CC infections or infections in wounds. In addition, the C5a antibody can

XX be used for treating sepsis. C5a is also known as anaphylatoxin.

PS Example 7; Page 26; 84pp; English.

XX The present sequence is human complement component C5a. The present

CC invention relates to an antibody specific for the present sequence. The

CC C5a-antibody can be used in a therapeutic composition, which is useful

CC for treating a subject suffering from bacterial infection, e.g.

CC sinusitis, meningitis, respiratory, gastrointestinal or urinary tract

CC infections or infections in wounds. In addition, the C5a antibody can

RESULT 9
AAR77604
ID AAR77604 standard; protein: 1676 AA.
AC AAR77604;
XX
DT 15-MAR-1996 (first entry)
XX
DE Pro-C5 polypeptide.
XX
KM Complement C5: haemolysis; kidney; glomerulonephritis;
KM monoclonal antibody; antinflammatory; antibody engineering;
KW humanised antibody.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..18
FT /label= Sig-peptide
FT Protein 19..673
FT /label= Beta-chain
FT Cleavage-site 673..674
FT Cleavage-site 677..678
FT Peptide 674..677
FT /label= Cleavage-peptide
FT Protein 678..1676
FT /label= Alpha-chain
FT /note= "amino acids 872-892 (854-874 of
FT the mature protein) comprise the KSSKS
FT epitope"
FT Peptide 678..751
FT /label= C5a
FT Cleavage-site 751..752
FT /label= Convertase_cleavage-site
FT Modified-site 911
FT /label= N-glycosylation-site
FT Modified-site 1115
FT /label= N-glycosylation-site
FT Modified-site 1630
FT /label= N-glycosylation-site
XX
PN W09529697-A1.
XX
PD 09-NOV-1995.
XX
PF 01-MAY-1995; 95WO-US05688.
XX
PR 02-MAY-1994; 94US-0236208.
XX
PA (ALEX-) ALEXION PHARM INC.
XX
PI Evans MJ, Matlis L, Mueller BE, Nye SH, Rollins S;
PI Rother RP, Springhorn J P, Squinto SP, Thomas TC;
PI Wang Y, Wilkins JA.
XX
DR WPI: 1995-392923/50.
XX
PT Treating glomerulonephritis with antibody against complement C5
PT component - to inhibit complement induced cell lysis
XX
PS Example 13: Page 82-92; 181pp; English.
XX
CC The cDNA sequence of the complement C5 gene transcript predicts a
CC secreted pro-C5 precursor of 1676 amino acids (AAR77604). C5 is a
CC beta-globulin heterodimer thought to play a role in the pathogenesis
CC of glomerulonephritis (GN). Cleavage of the C5 alpha-chain
CC by a convertase enzyme generates anaphylatoxic C5a. Monoclonal
CC and humanised recombinant antibodies that recognise the alpha-chain
CC KSSKC epitope (AAR77605) block C5a generation, thereby reducing
CC glomerular inflammation and kidney dysfunction associated with GN.
XX
SQ Sequence 1676 AA:

Query Match 100.0%; Score 52; DB 16; Length 1676;
Best Local Similarity 100.0%; Pred. No. 0.052;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ISHKMQLGR 10
DB 742 ISHKMQLGR 751
RESULT 10
AAR75501
ID AAR75501 standard; Peptide: 10 AA.
XX
AC AAR75501;
XX
DT 13-NOV-1995 (first entry)
XX
DE Control decapeptide.
XX
NE C5a: C5a receptor-antagonist; antinflammatory.
XX
KW Synthetic.
XX
PN W09516033-A.
XX
PD 15-JUN-1995.
XX
PF 16-NOV-1994; 94WO-1800359.
XX
PR 06-DEC-1993; 93US-0162591.
XX
PA (CIBA) CIBA GEIGY AG.
XX
PI Boyar WC, Galakatos NC, Peppard JV, Van Oostrum J;
XX
DR WPI: 1995-224319/29.
XX
PT C5a receptor antagonists having no agonist activity - are used in
PT compns. to treat C5a-mediated diseases and inflammatory conditions
XX
PS Example 10: Page 29; 65pp; English.
XX
CC C5a encoded by a synthetic gene is given in AAR75497. Analogs of C5a,
CC obd. by mutagenesis of the C-terminal region (AAR75500), exhibit
CC excellent antagonistic properties and substantially no agonist
CC activity. Preferred analogs are C5a(1-72), C5a(1-71), C5a(1-70),
CC C5a(1-70), C5a(1-69), C5a(1-68), C5a(1-67), C5a(1-66), C5a(1-65),
CC C5a(1-64), C5a(1-63), C5a(1-62), C5a(1-61), C5a(1-60), C5a(1-59),
CC C5a(1-58), C5a(1-57), C5a(1-56), C5a(1-55), C5a(1-54), C5a(1-53),
CC C5a(1-52), C5a(1-51), C5a(1-50), C5a(1-49), C5a(1-48), C5a(1-47),
CC C5a(1-46), C5a(1-45), C5a(1-44), C5a(1-43), C5a(1-42), C5a(1-41),
CC C5a(1-40), C5a(1-39), C5a(1-38), C5a(1-37), C5a(1-36), C5a(1-35),
CC C5a(1-34), C5a(1-33), C5a(1-32), C5a(1-31), C5a(1-30), C5a(1-29),
CC C5a(1-28), C5a(1-27), C5a(1-26), C5a(1-25), C5a(1-24), C5a(1-23),
CC C5a(1-22), C5a(1-21), C5a(1-20), C5a(1-19), C5a(1-18), C5a(1-17),
CC C5a(1-16), C5a(1-15), C5a(1-14), C5a(1-13), C5a(1-12), C5a(1-11),
CC C5a(1-10), C5a(1-9), C5a(1-8), C5a(1-7), C5a(1-6), C5a(1-5), C5a(1-4),
CC C5a(1-3), C5a(1-2), C5a(1-1), C5a(1-0).
XX
SQ Sequence 10 AA:
Query Match 82.7%; Score 43; DB 16; Length 10;
Best Local Similarity 90.0%; Pred. No. 0.012;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ISHKMQLGR 10
DB 1 ISHKMQLGR 10
RESULT 11
AAR94464
ID AAR94464 standard; peptide: 10 AA.
XX
AC AAR94464;
XX
DT 17-OCT-1996 (first entry)
XX
DE Human C5a anaphylatoxin C-terminal region comprising residues 65-74.

XX C-terminal: human: C5a anaphylatoxin; analogue: beta-turn:
 KW constrained backbone conformation; proinflammatory response;
 KW spasmogenesis; platelet aggregation; neutrophil: non-mediated;
 KW mediated: increase: cell membrane: vascular: permeability;
 KW neutrophil polarisation; neutrophil enzyme release; treatment;
 KW immune adjuvant; immunodeficiency; augmentation; immune therapy;
 KW cancer; drug delivery; CNS disorder; central nervous system;
 KW Alzheimer's disease; tumour: high affinity: receptor antagonist;
 KW development; non-steroidal; anti-inflammatory.

OS Homo sapiens.
 XX WO9606629-A1.
 XX 07-MAR-1996.
 XX PD
 XX 31-AUG-1995: 95WO-US11126.
 XX PR 31-AUG-1994: 94US-0299285.
 XX PA (UYNE-) UNIV NEBRASKA.
 XX PA (UYQU) UNIV QUEENSLAND.
 XX PI Klnarsky L, Sanderson DG, Sherman SA, Taylor SM;
 XX WPI: 1996-160140/16.
 XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
 PT spasmogenesis, platelet aggregation and increases in cell membrane
 PT permeability

Example 2: Page 77: 116pp: English.

The present peptide is a C-terminal decapeptide region from human C5a anaphylatoxin, which has a flexible C-terminal region. Using this sequence as a template, C5a analogues with a constrained backbone conformation comprising a type (II or V) or type III beta-turn, were prepd. The type (II or V) and type III analogues elicit proinflammatory responses characterised by spasmogenesis, platelet aggregation and neutrophil non-mediated increases of cell membrane (specifically vascular) permeability, or neutrophil polarisation, neutrophil enzyme release and neutrophil mediated increases in cell membrane (esp. vascular) permeability, respectively.

The analogues can be used as immune adjuvants for the treatment of immune deficiency disorders, and for augmenting standard immune therapy for the treatment of cancer, without inflammatory side effects. They can also be used to facilitate drug delivery by increasing vascular permeability, e.g. in the treatment of CNS disorders such as Alzheimer's disease or tumours, and to develop high affinity C5a receptor antagonists useful as non-steroidal anti-inflammatory agents.

In a type (II or V) beta-turn specific guinea pig platelet aggregation assay, and a type III beta-turn specific neutrophil polarisation assay the peptide had respective (microm) and EC50 (microm) activities of 0.023 and 0.09, compared to 0.023 and 0.0013 for C5a.

Sequence 10 AA:

Query Match 82.7%; Score 43; DB 17; Length 10;
 Best Local Similarity 90.0%; Pred. No. 0.012;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ISHKMOQLR 10
 1111111111
 Db 1 isfkdmqlgr 10

ID AAR94465 standard; peptide: 10 AA.
 XX AAR94465:
 AC 17-OCT-1996 (first entry)
 XX
 DT C5a anaphylatoxin C-terminal analogue C5a(65-74) Tyr65.
 XX
 XX C-terminal: human: C5a anaphylatoxin; analogue: beta-turn:
 KW constrained backbone conformation; proinflammatory response;
 KW spasmogenesis; platelet aggregation; neutrophil: non-mediated;
 KW mediated: increase: cell membrane: vascular: permeability;
 KW neutrophil polarisation; neutrophil enzyme release; treatment;
 KW immune adjuvant; immunodeficiency; augmentation; immune therapy;
 KW cancer; drug delivery; CNS disorder; central nervous system;
 KW Alzheimer's disease; tumour: high affinity: receptor antagonist;
 KW development; non-steroidal; anti-inflammatory.

OS Homo sapiens.
 XX WO9606629-A1.
 XX 07-MAR-1996.
 XX PD
 XX 31-AUG-1995: 95WO-US11126.
 XX PR 31-AUG-1994: 94US-0299285.
 XX PA (UYNE-) UNIV NEBRASKA.
 XX PA (UYQU) UNIV QUEENSLAND.
 XX PI Klnarsky L, Sanderson DG, Sherman SA, Taylor SM;
 XX WPI: 1996-160140/16.
 XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
 PT spasmogenesis, platelet aggregation and increases in cell membrane
 PT permeability

Claim 8: Page 77: 116pp: English.

The present peptide is an analogue of a C-terminal decapeptide region from human C5a anaphylatoxin, which has a constrained backbone conformation (compared to the natural peptide's flexible C-terminal region) comprising a beta-turn. Analogues with type (II or V) and type III beta-turns elicit proinflammatory responses characterised by spasmogenesis, platelet aggregation and neutrophil non-mediated increases of cell membrane (specifically vascular) permeability, or neutrophil polarisation, neutrophil enzyme release and neutrophil mediated increases in cell membrane (esp. vascular) permeability, respectively.

The analogues can be used as immune adjuvants for the treatment of immune deficiency disorders, and for augmenting standard immune therapy for the treatment of cancer, without inflammatory side effects. They can also be used to facilitate drug delivery by increasing vascular permeability, e.g. in the treatment of CNS disorders such as Alzheimer's disease or tumours, and to develop high affinity C5a receptor antagonists useful as non-steroidal anti-inflammatory agents.

In a type (II or V) beta-turn specific human foetal artery smooth muscle contraction assay, and a type III beta-turn specific neutrophil polarisation assay the peptide had respective EC50 (microm) activities of 11.2 and 4.0, compared to 0.018 and 0.0013 for C5a.

Sequence 10 AA:

Query Match 75.0%; Score 39; DB 17; Length 10;

Best Local Similarity 88.9%; Pred. No. 0.085;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 SHKMOGLR 10
1111111111
Db 2 stkmqlgr 10

RESULT 13

AAE05455
ID AAE05455 standard: Protein: 74 AA.

XX AAE05455;

XX 24-SEP-2001 (first entry)

DE Pig C5a anaphylatoxin.

XX Mannan binding protein-associated serine protease; MASP;

KM complement-activation; C-activation; microbial infection;

KM cytomagalovirus; CMV; hepatitis virus; human immunodeficiency virus;

KM HIV; organ transplant rejection; tissue injury; autoimmune disease;

KM rheumatoid arthritis; systemic lupus erythematosus; SLE; pig;

KM inflammatory response; Alzheimer's disease; C5a anaphylatoxin.

XX Sus scrofa.

XX Key Location/Qualifiers

FT Modified-site 64..66

FT /note= "Asn is N-glycosylated"

XX US6235494-B1.

XX 22-MAY-2001.

XX 08-FEB-1999; 99US-0246500.

XX 08-FEB-1999; 99US-0246500.

XX (SCRI) SCRIPPS RES INST.

XX Hugli TE;

XX WPI: 2001-450082/48.

XX Claim 3; Column 49-50; 39pp; English.

XX The invention relates to assays for measuring in-vivo levels of

XX activated mannan-binding protein-associated serine protease (MASP-1 and

XX MASP-2) activity. The assay comprises contacting a sample of blood or

XX plasma comprising a metal ion chelator with a substrate of formula:

XX R-peptide-Y; wherein the peptide comprises at least 4 residues from the

XX C-terminus of an anaphylatoxin selected from C3a, C4a and C5a; R is a

XX peptide blocking, protecting or capping group; and Y is any group

XX cleavable from the substrate by MASP and comprises a labelled tag. The

XX substrate of the invention is also useful for monitoring in vitro and

XX in vivo complement-activation (C-activation) by classical, alternative

XX or lectin pathways. The assays of MASP activity can be used for

XX detecting or monitoring a condition associated with complement

XX activation. The conditions include microbial infections

XX caused by particularly cytomegalovirus (CMV), hepatitis virus and

XX human immunodeficiency virus (HIV), organ transplant rejection, tissue

XX injury, autoimmune diseases (e.g. rheumatoid arthritis, systemic lupus

XX erythematosus (SLE)) and inflammatory responses (e.g. in Alzheimer's

XX disease and bacterial diseases). The assay of MASP is also useful for

XX assessing the toxicity or injury of therapeutic treatments or screening

XX test compounds as agents for treatment of viral diseases, parasitic

XX infections, tissue injury, organ transplant rejection, autoimmune

XX diseases or inflammatory responses. The present sequence is pig

CC C5a anaphylatoxin.
XX
SQ Sequence 74 AA;

Query Match 75.0%; Score 39; DB 22; Length 74;
Best Local Similarity 77.8%; Pred. No. 0.81;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 SHKMOGLR 10
1111111111
Db 66 shkniqlgr 74

RESULT 14

AAB74057
ID AAB74057 standard: protein: 74 AA.

XX AAB74057;

DT 16-MAY-2001 (first entry)

DE Bovine C5a.

KM Bovine; C5a; complement; antibody; bacterial infection; sinusitis;

KM meningitis; respiratory; gastrointestinal; urinary tract infection;

KM wound; anaphylatoxin; sepsis.

XX Bos sp.

XX WO200115731-A1.

XX 31-AUG-2000; 2000WO-US24219.

XX 31-AUG-1999; 99US-0387671.

XX (UNMI) UNIV MICHIGAN.

XX Ward PA, Huber-Lang M, Sarma V;

XX WPI: 2001-226665/23.

XX Compositions for treating blood-borne and toxin mediated diseases and

XX treatment of sepsis in humans and other animals comprises anti-C5a

XX antibodies generated against C-terminal truncated C5a peptides

XX Disclosure: Page 26; 84pp; English.

XX The present sequence is bovine complement component C5a. The present

XX invention relates to an antibody specific for the present sequence. The

XX C5a-antibody can be used in a therapeutic composition, which is useful

XX for treating a subject suffering from bacterial infection, e.g.

XX sinusitis, meningitis, respiratory, gastrointestinal or urinary tract

XX infections or infections in wounds. In addition, the C5a antibody can

XX be used for treating sepsis. C5a is also known as anaphylatoxin.

XX Sequence 74 AA;

OY 3 HKDMOGLR 10
1111111111
Db 67 hkmqlgr 74

RESULT 15

AAB74058
ID AAB74058 standard: protein: 74 AA.

Query Match 75.0%; Score 39; DB 22; Length 74;

Best Local Similarity 87.5%; Pred. No. 0.81; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

XX AAB74058;
AC
XX
XX

DT 16-MAY-2001 (first entry)
XX
XX

DE Porcine C5a.

XX Porcine; C5a; complement; antibody; bacterial infection; sinusitis;
KW meningitis; respiratory; gastrointestinal; urinary tract infection;
KM wound; anaphylatoxin; sepsis.
XX

OS Sus scrofa.

XX WO200115731-A1.
PN
XX

PD 08-MAR-2001.
XX

PF 31-AUG-2000; 2000MO-US24219.
XX

PR 31-AUG-1999; 99US-0387671.
XX

PA (UNMI) UNIV MICHIGAN.
XX

PI Ward PA, Huber-Lang M, Sarma V;
XX

DR WPI: 2001-22665/23.
XX

XX Compositions for treating blood-borne and toxin mediated diseases and
PT treatment of sepsis in humans and other animals comprises anti-C5a
PT antibodies generated against C-terminal truncated C5a peptides -
XX

PS Disclosure; Page 26; 84pp; English.
XX

XX The present sequence is porcine complement component C5a. The present
CC invention relates to an antibody specific for the present sequence. The
CC C5a-antibody can be used in a therapeutic composition, which is useful
CC for treating a subject suffering from bacterial infection, e.g.
CC sinusitis, meningitis, respiratory, gastrointestinal or urinary tract
CC infections or infections in wounds. In addition, the C5a antibody can
CC be used for treating sepsis. C5a is also known as anaphylatoxin.
XX

SQ Sequence 74 AA;

Query Match 75.0%; Score 39; DB 22; Length 74;

Best Local Similarity 77.8%; Pred. No. 0.81;

Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 SHKDMOLGR 10
|||:||||

DB 66 shkniqlgr 74

Search completed: February 27, 2002, 11:41:12
Job time: 451 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:33:41 ; Search time 145.23 Seconds

(without alignments)
5.245 Million cell updates/sec

Title: us-09-446-109a-1

Perfect score: 52

Sequence: 1 ISHKDQLGR 10

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	100.0	1676	1 CSHU	complement C5 prec
2	41	78.8	306	2 S41291	ornithine carbamoy
3	39	75.0	74	2 A25408	complement C5 - bo
4	39	75.0	74	2 A01268	complement C5 - pl
5	39	75.0	319	1 OWBS	ornithine carbamoy
6	38	73.1	1149	2 S67099	DNA-directed RNA p
7	37	71.2	715	2 T25233	hypothetical prote
8	35	67.3	278	2 S36387	probable protein k
9	35	67.3	319	2 F84011	ornithine carbamoy
10	35	67.3	1489	2 G83771	hypothetical prote
11	34	65.4	77	2 A57689	complement C5a - r
12	34	65.4	220	2 T09108	RNA binding protei
13	34	65.4	225	2 S38324	peptidylprolyl iso
14	34	65.4	363	2 A72702	hypothetical prote
15	34	65.4	408	1 E42409	biphenyl dioxygena
16	34	65.4	408	1 F41858	biphenyl dioxygena
17	34	65.4	547	2 S44841	K06H7.1 protein -
18	34	65.4	591	2 T39195	probable amino aci
19	34	65.4	648	2 T43337	polo-like kinase-1
20	34	65.4	743	2 E83728	phosphoribosylform
21	34	65.4	1541	1 S71839	canalicular multid
22	33	63.5	129	2 T26142	hypothetical prote
23	33	63.5	291	2 B64629	hypothetical prote
24	33	63.5	292	2 B71885	hypothetical prote
25	33	63.5	293	2 B72325	riboflavin kinase/
26	33	63.5	1508	2 T27828	hypothetical prote
27	33	63.5	1519	2 T27829	hypothetical prote
28	33	63.5	1680	1 C5MS	complement C5 prec
29	33	63.5	3187	2 JC5837	36k Golgi complex

30	32	61.5	116	2 B64451	hypothetical prote
31	32	61.5	136	1 HS2P3	histone H3.1 [simi
32	32	61.5	137	2 S37870	chromatin-associat
33	32	61.5	256	2 S07105	actin-binding prot
34	32	61.5	284	2 G01229	capring protein al
35	32	61.5	286	2 A33546	actin-capping prot
36	32	61.5	286	2 S36093	actin-capping prot
37	32	61.5	286	2 G02639	capring protein al
38	32	61.5	295	2 A84096	hypothetical prote
39	32	61.5	387	2 E69059	ornithine carbamoy
40	32	61.5	387	2 C66072	hypothetical prote
41	32	61.5	416	2 E64140	hypothetical prote
42	32	61.5	462	1 S00552	mitochondrial proc
43	32	61.5	574	2 T29137	hypothetical prote
44	32	61.5	596	2 A96539	hypothetical prote
45	32	61.5	690	2 E84724	hypothetical prote

ALIGNMENTS

RESULT 1
CSHU
Complement C5 precursor [validated] - human
N:Contains: C5a anaphylatoxin: C5b
C:Species: Homo sapiens (man)
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text-change 08-Dec-2000
C:Accession: A40075; A27689; A01267; A01266; S15121
C:Raviland, D.L.; Haviland, J.C.; Fleischner, D.T.; Hunt, A.; Wetzel, R.A.
J. Immunol. 146, 362-368, 1991
A:Title: Complete cDNA sequence of human complement pro-C5. Evidence of truncated tra
A:Reference number: A40075; MUID:91079575
A:Accession: A40075
A:Molecule type: mRNA
A:Residues: 1-1676 <HAV>
A:Cross-references: CB:M57729; MID:q179982; PIDN:AAA51925.1; PID:q179983
A:Note: 518-Ser was also found
A:Wetzel, R.A.; Lemons, R.S.; Le Beau, M.M.; Barnum, S.R.; Noack, D.; Tack, B.F.
Biochemistry 27, 1474-1482, 1988
A:Title: Molecular analysis of human complement component C5: Localization of the str
A:Reference number: A27689; MUID:88209511
A:Accession: A27689
A:Molecule type: mRNA
A:Residues: 412-1676 <WET>
A:Cross-references: CB:M65134; CB:M18879; MID:q179691; PIDN:AAA51856.1; PID:q179692
A:Raviland, D.L.; Haviland, J.C.; Fleischner, D.T.; Hunt, A.; Wetzel, R.A.
J. Biol. Chem. 253, 6955-6964, 1978
A:Title: Primary structural analysis of the polypeptide portion of human C5a anaphyla
A:Reference number: A01267; MUID:79005687
A:Accession: A01267
A:Molecule type: Protein
A:Residues: 678-731 <FER>
A:Raviland, D.L.; Haviland, J.C.; Fleischner, D.T.; Hunt, A.; Wetzel, R.A.
J. Biol. Chem. 260, 2108-2112, 1985
A:Title: Isolation and sequence analysis of a cDNA clone encoding the fifth complemen
A:Reference number: A01266; MUID:85130937
A:Accession: A01266
A:Molecule type: mRNA
A:Residues: 412-854; 'SLA5PRLECNKISGNCKLRPGSSDSPAASQVAGITGNNNAQPT' <LUN>
A:Cross-references: CB:K02874
A:Note: The carboxyl-terminal part of the sequence in this report appears to be deriv
A:Bohnsack, J.F.; Mollison, K.W.; Buko, A.M.; Ashworth, J.C.; Hill, H.R.
Biochem. J. 273, 635-640, 1991
A:Title: Group B streptococci inactivate complement component C5a by enzymic cleavage
A:Reference number: S15121; MUID:91144547
A:Accession: S15121
A:Comment: complement C5 contains two disulfide-linked chains, formed by removal of I
(beta and alpha' chains).
C:Comment: Activation of C5 initiates the spontaneous assembly of the late complement
is the foundation upon which the membrane attack complex is assembled.
C:Comment: C5a has potent spasmogenic and chemotactic activity.
C:Genetics:
A:Gene: GDB:C5

A:Cross-references: GDB:119734; OMIM:120900
 A:Map position: 9q34.1-9q34.1
 C:Superfamily: alpha-2-macroglobulin
 C:Keywords: complement alternate pathway; complement pathway; cytolytic; glycoprotein; 1
 F:1-18/Domain: signal sequence #status predicted <SIG>
 F:19-673/678-1676/Product: complement C5 #status predicted <KAT>
 F:19-673/752-1676/Product: C5b #status predicted <CSB>
 F:19-673/Product: complement C5 and C5b beta chain #status predicted <CSB>
 F:678-1676/Product: complement C5 alpha chain #status predicted <CSA>
 F:678-1676/Product: C5a anaphylatoxin #status experimental <C5T>
 F:752-1676/Product: C5b alpha' chain #status predicted <CSBA>
 F:567-810/634-669/724/699-731/711-732/866-1527/1101-1159/1375-1505/1405-1474/1520-15
 F:741/Binding: site: carbohydrate (Asn) (covalent) #status experimental
 F:751-752/Cleavage site: Arg-Leu (C5 convertase) #status experimental
 F:911/1115/1630/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0% Score 52: DB 1: Length 1676;
 Best Local Similarity 100.0% Pred. No. 0.022;
 Matches 10: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

OY 1 ISHKMOLGR 10
 |||:|||||
 Db 742 ISHKMOLGR 751

RESULT 2

ornithine carbamoyltransferase (EC 2.1.3.3) - Pseudomonas syringae
 C:Species: Pseudomonas syringae
 C:Date: 19-May-1994 #sequence_revision 03-Aug-1995 #text_change 18-Jun-1999
 C:Accession: S41291
 R:Mosqueda-Cano, G.
 Submitted to the EMBL Data Library, December 1993
 A:Reference number: S41291
 A:Accession: S41291
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-306 <MOS>
 A:Cross-references: EMBL:X76945; NID:9440573; PIDN:CA54264.1; PID:9440574
 C:Superfamily: ornithine carbamoyltransferase; aspartate/ornithine carbamoyltransferase
 C:Keywords: transfease
 F:4-299/Domain: aspartate/ornithine carbamoyltransferase homology <ACT>

Query Match 78.8% Score 41: DB 2: Length 306;
 Best Local Similarity 70.0% Pred. No. 0.71;
 Matches 7: Conservative 2: Mismatches 1: Indels 0: Gaps 0:

OY 1 ISHKMOLGR 10
 |||:|||||
 Db 74 LSHKDTOLGR 83

RESULT 3

complement C5 - bovine (fragment)
 N:Contains: C5a anaphylatoxin; C5b
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 02-Jun-1988 #sequence_revision 15-Oct-1994 #text_change 16-Feb-1997
 C:Accession: A25408
 R:Gennaro, R.; Simonic, T.; Negrl, A.; Mottola, C.; Secchi, C.; Ronchi, S.; Romeo, D.
 Eur. J. Biochem. 155: 77-86, 1986
 A:Title: C5a fragment of bovine complement. Purification, bioassays, amino-acid sequence
 A:Reference number: A25408; MUID:86136134
 A:Accession: A25408
 A:Molecule type: protein
 A:Residues: 1-74 <GEN>
 C:Comment: Complement C5 contains two disulfide-linked chains, formed by removal of four
 (beta and alpha' chains).
 C:Comment: Activation of C5 initiates the spontaneous assembly of the late complement co
 is the foundation upon which the membrane attack complex is assembled.
 C:Comment: C5a has potent spasmogenic and chemotactic activity.

C:Superfamily: alpha-2-macroglobulin
 C:Keywords: complement alternate pathway; complement pathway; cytolytic; glycoprotein
 F:1-74/Product: C5a anaphylatoxin #status experimental <C5T>
 F:21-47/22-54/34-55/Disulfide bonds: #status predicted

Query Match 75.0% Score 39: DB 2: Length 74;
 Best Local Similarity 87.5% Pred. No. 0.39;
 Matches 7: Conservative 1: Mismatches 0: Indels 0: Gaps 0:

OY 3 HKDMOLGR 10
 |||:|||||
 Db 67 HKDMOLGR 74

RESULT 4

complement C5 - pig (fragment)
 N:Contains: C5a anaphylatoxin; C5b
 C:Species: Sus scrofa domestica (domestic pig)
 C:Date: 29-Jul-1981 #sequence_revision 15-Oct-1994 #text_change 16-Feb-1997
 C:Accession: A01268; A26248
 R:Gerard, C.; Hugli, T.E.
 J. Biol. Chem. 255: 4710-4715, 1980
 A:Title: Amino acid sequence of the anaphylatoxin from the fifth component of porcine
 A:Reference number: A01268; MUID:80182137
 A:Accession: A01268
 A:Molecule type: protein
 A:Residues: 1-74 <GER>
 R:Gerard, C.; Hugli, T.E.
 Proc. Natl. Acad. Sci. U.S.A. 78: 1833-1837, 1981
 A:Reference number: A26247; MUID:81199549
 A:Contents: annotation: active region
 A:Note: although Arg-74 is not essential, residues 72-74 (Leu-Gly-Arg) are required f

A:Reference number: A26248; MUID:84184201
 A:Contents: disulfide bonds
 A:Accession: A26248
 A:Molecule type: protein
 A:Residues: 1-64, E', 66-73 <2IM>
 C:Comment: Complement C5 contains two disulfide-linked chains, formed by removal of f

(beta and alpha' chains).
 C:Comment: Activation of C5 initiates the spontaneous assembly of the late complement
 is the foundation upon which the membrane attack complex is assembled.
 C:Comment: C5a has potent spasmogenic and chemotactic activity.
 C:Superfamily: alpha-2-macroglobulin
 C:Keywords: complement alternate pathway; complement pathway; cytolytic; glycoprotein
 F:1-74/Product: C5a anaphylatoxin #status experimental <C5T>
 F:21-47/22-54/34-55/Disulfide bonds: #status experimental

Query Match 75.0% Score 39: DB 2: Length 74;
 Best Local Similarity 77.8% Pred. No. 0.39;
 Matches 7: Conservative 2: Mismatches 0: Indels 0: Gaps 0:

OY 2 SHKMOLGR 10
 |||:|||||
 Db 66 SHKMOLGR 74

RESULT 5

ornithine carbamoyltransferase (EC 2.1.3.3) - Bacillus subtilis
 N:Alternate names: citrulline phosphorilase; ornithine transcarbamylase
 C:Species: Bacillus subtilis
 C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Jun-2000
 C:Accession: S11000; A38768; I40378; A69589; S38434
 R:Mountain, A.; Smith, M.C.M.; Baumberg, S.
 Nucleic Acids Res. 18: 4594, 1990
 A:Title: Nucleotide sequence of the Bacillus subtilis argF gene encoding ornithine ca
 A:Reference number: S10999; MUID:90356402
 A:Accession: S11000

A:Molecule type: DNA
 A:Residues: 1-319 <MOU>
 A:Cross-references: EMBL:X53360; NID:q39809; PIDN:CAA37444.1; PID:q39811
 A:Accession: A38768
 A:Molecule type: protein
 A:Residues: 1,'X',3-14,'XX',17-20,'X',22-30,'XX',33,'X',35-39 <MOU2>
 R:O'Reilly, M.; Devine, K.M.
 Microbiology 140, 1023-1025, 1994
 A:Title: Sequence and analysis of the citrulline biosynthetic operon argC-F from *Bacillus*
 A:Reference number: 140372; MUID:94297722
 A:Accession: 140378
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-319 <RES>
 A:Cross-references: EMBL:Z26919; NID:q408113; PIDN:CAA81542.1; PID:q408120
 R:Kunst, F.; Ogawa, N.; Moser, I.; Albertin, A.M.; Alloni, G.; Azevedo, V.; Berter
 C.; Bron, S.; Broillet, S.; Brusch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.; Chd
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrati, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Gallizzi, A.; Gall
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.
 Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
 Y.; M.; Ogawa, K.; Ogilwa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portelle
 Rieger, M.; Rivoita, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot
 akuchl, M.; Tamashiro, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
 A:Reference number: A69580; MUID:98044033
 A:Accession: A69589
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-319 <KUN>
 A:Cross-references: GB:Z99109; GB:Z99110; GB:AL009126; NID:q2633472; PIDN:CAB12982.1; PI
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: argF
 A:Map position: 100 (degrees)
 C:Superfamily: ornithine carbamoyltransferase: aspartate/ornithine carbamoyltransferase
 C:Keywords: arginine biosynthesis; transferase
 F:1-319/Product: ornithine carbamoyltransferase \$status experimental <MAT>
 F:12-310/Domain: aspartate/ornithine carbamoyltransferase homology <ACT>

Query Match 75.0%; Score 39; DB 1; Length 319;
 Best Local Similarity 70.0%; Pred. No. 2;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISHKMQLGR 10
 DB 78 LSQKDLQGR 87

RESULT 6
 S67099
 DNA-directed RNA polymerase (EC 2.7.7.6) III second-largest chain - yeast (*Saccharomyces*
 N:Alternate names: protein Q846; protein YOR207C
 C:Species: *Saccharomyces cerevisiae*
 C:Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 20-Jun-2000
 C:Accession: S67099; S14169
 R:Hughes, B.; Pohl, T.M.
 submitted to the Protein Sequence Database, July 1996
 A:Reference number: S66685
 A:Accession: S67099
 A:Molecule type: DNA
 A:Residues: 1-1149 <HUC>
 A:Cross-references: EMBL:Z75115; NID:q1420484; PIDN:CAA99422.1; PID:q1420485; MIPS:YOR20
 A:Experimental source: strain S288C
 R:James, P.; Whelen, S.; Hall, B.D.
 J. Biol. Chem. 266, 5616-5624, 1991
 A:Title: The RRT1 gene of yeast encodes the second-largest subunit of RNA polymerase III

A:Reference number: S14169; MUID:91170230
 A:Accession: S14169
 A:Molecule type: DNA
 A:Residues: 1'212','E',214-1149 <JBT>
 A:Cross-references: GB:M58723; NID:q172381; PIDN:AAB59324.1; PID:q172383
 C:Genetics:
 A:Gene: SGD:RRT1; RPC128; RPC2
 A:Cross-references: SGD:S0005733; MIPS:YOR207C
 A:Map position: 15R
 C:Superfamily: DNA-directed RNA polymerase 132K polypeptide
 C:Keywords: nucleotidyltransferase; transcription

Query Match 73.1%; Score 38; DB 2; Length 1149;
 Best Local Similarity 60.0%; Pred. No. 13;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISHKMQLGR 10
 DB 142 IMHKDVELGR 151

RESULT 7
 T25233
 Hypothetical protein T24D1.1 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T25233
 R:Cummings, P.
 submitted to the EMBL Data Library, October 1996
 A:Reference number: Z20001
 A:Accession: T25233
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-715 <MIL>
 A:Cross-references: EMBL:Z81131; PIDN:CAB03422.1; GSPDB:GN00019; CESP:T24D1.1
 A:Experimental source: clone T24D1
 C:Genetics:
 A:Gene: CESP:T24D1.1
 A:Map position: 1
 A:Intons: 3/3; 99/3; 210/2; 371/3; 514/2; 629/3; 655/3

Query Match 71.2%; Score 37; DB 2; Length 715;
 Best Local Similarity 66.7%; Pred. No. 13;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 SHKDMQLGR 10
 DB 226 SHEDVELGR 234

RESULT 8
 S36387
 Probable protein kinase (EC 2.7.1.37) cdc2/cdc28-related - *Trichomonas vaginalis*
 C:Species: *Trichomonas vaginalis*
 C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Dec-1997
 C:Accession: S36387; S32884
 R:Riley, D.E.; Campbell, L.; Puolakkainen, M.H.; Krieger, J.N.
 submitted to the EMBL Data Library, February 1993
 A:Reference number: S36387
 A:Accession: S36387
 A:Molecule type: DNA
 A:Residues: 1-278 <RLD>
 A:Cross-references: EMBL:L10131
 R:Riley, D.E.; Campbell, L.A.; Puolakkainen, M.; Krieger, J.N.
 Mol. Microbiol. 8, 517-519, 1993
 A:Title: *Trichomonas vaginalis* and early evolving DNA and protein sequences of the CD
 A:Reference number: S32884; MUID:93316848
 A:Accession: S32884
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 11-146 <RL2>

A:Cross-references: EMBL:L10131
 C:Superfamily: kinase-related transforming protein; protein kinase homology
 C:Keywords: Atp; phosphotransferase; protein kinase
 F:4-253/Domain: protein kinase homology <kin>
 F:12-20/Region: protein kinase ATP-binding motif

Query Match 67.3%; Score 35; DB 2; Length 278;
 Best Local Similarity 85.7%; Pred. No. 12;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 HKDMQLGR 9
 |||:|:
 Db 7 HKDKMLG 13

RESULT 9
 F84011
 ornithine carbamoyltransferase argf [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence-revision 01-Dec-2000 #text-change 08-Dec-2000
 C:Accession: F84011
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai
 Nucleic Acids Res. 28: 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
 A:Reference number: A83650; MUID:20263314
 A:Accession: F84011
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-319 <STO>
 A:Cross-references: GB:AP001517; GB:BA000004; MID:g10175500; PIDN:BA06613.1; GSPDB:GNOC
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: argf
 C:Superfamily: ornithine carbamoyltransferase; aspartate/ornithine carbamoyltransferase

Query Match 67.3%; Score 35; DB 2; Length 319;
 Best Local Similarity 60.0%; Pred. No. 14;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 ISHKMQLGR 10
 :|||:|:
 Db 84 LSPKQLQIGR 93

RESULT 10
 C83771
 hypothetical protein BH0975 [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence-revision 01-Dec-2000 #text-change 31-Dec-2000
 C:Accession: C83771
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai
 Nucleic Acids Res. 28: 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
 A:Reference number: A83650; MUID:20263314
 A:Accession: C83771
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1489 <STO>
 A:Cross-references: GB:AP001510; GB:BA000004; MID:g10173440; PIDN:BA064694.1; GSPDB:GNOC
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH0975

Query Match 67.3%; Score 35; DB 2; Length 1489;
 Best Local Similarity 75.0%; Pred. No. 75;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 HKDMQLGR 10
 |||:|:
 Db 1417 HKDVSIGR 1424

RESULT 11
 A57689
 complement C5a - rat (fragment)

N:Contains: C5a anaphylatoxin; C5b
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 23-Feb-1996 #sequence-revision 31-Jan-1997 #text-change 16-Feb-1997
 C:Accession: A57689
 R:Cui, L.; Carney, D.F.; Hugli, T.E.
 Protein Sci. 3: 1169-1177, 1994
 A:Title: Primary structure and functional characterization of rat C5a: an anaphylatoxin
 A:Reference number: A57689; MUID:95078724
 A:Accession: A57689
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 1-77 <CUF>
 C:Complex: Complement C5 contains two disulfide-linked chains, formed by removal of f
 (beta and alpha' chains).
 C:Function:
 A:Description: activation of C5 initiates spontaneous assembly of the late complement
 is the foundation for assembly of the membrane attack complex
 A:Pathway: complement alternate pathway; complement attack pathway
 A:Note: C5a has potent spasmogenic and chemotactic activity
 C:Superfamily: complement alternate pathway; complement pathway; cytolysis; glycoprotein
 F:1-77/product: C5a anaphylatoxin #status experimental <C5P>
 F:24-50,25-57,37-58/disulfide bonds: #status predicted

Query Match 65.4%; Score 34; DB 2; Length 77;
 Best Local Similarity 77.8%; Pred. No. 4.6;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 SHKMQLGR 10
 ||| |||:
 Db 69 SHKMQLGR 77

RESULT 12
 T09108
 RNA binding protein, 24k, chloroplast - spinach (fragment)
 C:Species: Spinacia oleracea (spinach)
 C:Date: 11-Jun-1999 #sequence-revision 11-Jun-1999 #text-change 21-Jan-2000
 C:Accession: T09108
 R:Abrahamson, S.L.; Roell, M.K.; Schuster, G.; Grissam, W.
 submitted to the EMBL Data Library, August 1995
 A:Description: Isolation, cDNA cloning and characterization of a 24 kD ribonucleoprotein
 A:Reference number: Z16566
 A:Accession: T09108
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-220 <ABR>
 A:Cross-references: EMBL:U34742; MID:g1015369; PID:g1015370
 A:Experimental source: cv. Marathon
 C:Genetics:
 A:Genome: nuclear
 C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein
 F:39-106/Domain: ribonucleoprotein repeat homology <RRR>

Query Match 65.4%; Score 34; DB 2; Length 220;
 Best Local Similarity 77.8%; Pred. No. 15;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 SHKMQLGR 10
 ||| |||:
 Db 115 SHKMQLGR 123

RESULT 13
 S38324

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:33:41 Search time 78.39 Seconds
(without alignments)
4.677 Million cell updates/sec

Title: US-09-446-109A-1
Perfect score: 52
Sequence: 1 ISHKMQLGR 10

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	100.0	1676	1 COS_HUMAN	P01031 homo sapien
2	41	78.8	306	1 OTCA_PSESH	Q02047 pseudomonas
3	39	75.0	74	1 COSA_BOVIN	P12082 bos taurus
4	39	75.0	74	1 COSA_PIG	P01032 sus scrofa
5	39	75.0	319	1 OTCA_BACSU	P18186 bacillus su
6	38	73.1	1149	1 RPY2_YEAST	P22276 saccharomyc
7	34	65.4	76	1 COSA_RAT	P08650 ratius norv
8	34	65.4	225	1 CYPD_YEAST	P31737 burkholderi
9	34	65.4	408	1 BPHG_BURGE	P34509 caenorhabdi
10	34	65.4	547	1 YMX1_CAREL	Q9UL18 schizosacch
11	34	65.4	591	1 YFPA_SCHPO	Q9K157 bacillus ha
12	34	65.4	743	1 PURL_BACHD	O63120 ratius norv
13	34	65.4	1541	1 MRP2_RAT	P06684 mus musculu
14	33	63.5	1680	1 COS_MOUSE	O58608 methanococc
15	32	61.5	116	1 YC11_METJA	P09988 schizosacch
16	32	61.5	135	1 H33_SCHPO	P10651 schizosacch
17	32	61.5	135	1 H33_SCHPO	P10651 schizosacch
18	32	61.5	229	1 CSE4_YEAST	P36012 saccharomyc
19	32	61.5	256	1 CAPA_XENLA	P25229 xenopus lae
20	32	61.5	284	1 CAZ1_MOUSE	P47753 mus musculu
21	32	61.5	286	1 CAZ1_CHICK	P31327 gallus gall
22	32	61.5	286	1 CAZ1_HUMAN	P52907 homo sapien
23	32	61.5	286	1 CAZ2_CHICK	P28497 gallus gall
24	32	61.5	286	1 CAZ2_HUMAN	P47753 homo sapien
25	32	61.5	286	1 CAZ2_MOUSE	P47753 mus musculu
26	32	61.5	462	1 OTC_METTH	O37495 methanobact
27	32	61.5	462	1 MPPB_YEAST	P10507 saccharomyc
28	32	61.5	465	1 C1XG_HAERIN	P44458 h clixg pro
29	32	61.5	724	1 K6A1_MOUSE	P18653 mus musculu
30	32	61.5	735	1 K6A1_HUMAN	Q15418 homo sapien
31	32	61.5	756	1 K6A1_RAT	O63531 ratius norv
32	32	61.5	756	1 K6A1_YEAST	P53894 saccharomyc
33	32	61.5	880	1 VP2_ROTBR	P12472 bovine rota

34	32	61.5	881	1 VP2_ROTBU	P17462 bovine rota
35	32	61.5	881	1 VP2_ROTBU	P22672 simian 11 r
36	32	61.5	890	1 VP2_ROTBU	P11231 human rotav
37	31	59.6	97	1 Y04E_BPT4	P07079 bacterioph
38	31	59.6	124	1 T2DB_HUMAN	Q15543 homo sapien
39	31	59.6	134	1 H3_ENTHI	Q06196 entamoeba h
40	31	59.6	135	1 H31_TERYP	P15511 tetrahymena
41	31	59.6	135	1 H32_TERYP	P15512 tetrahymena
42	31	59.6	244	1 ERM2_STRAU	P02979 staphylococ
43	31	59.6	244	1 ERM3_STRAU	P13957 staphylococ
44	31	59.6	244	1 ERM4_STRAU	P13978 staphylococ
45	31	59.6	244	1 ERM5_STRAU	P06572 staphylococ

ALIGNMENTS

RESULT 1
ID COS_HUMAN STANDARD; PRT; 1676 AA.
AC P01031:
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE COMPLEMENT C5 PRECURSOR (CONTAINS: C5A ANAPHYLATOXIN).
GN C5.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=91079575; PubMed=1984448;
RA Haviland D.L., Haviland J.C., Fleischer D.T., Hunt A., Wetsel R.A.:
RT "Complete cDNA sequence of human complement pro-C5. Evidence of
truncated transcripts derived from a single copy gene.";
J. Immunol. 146:362-368(1991).
RN 12
RP SEQUENCE OF 412-1676 FROM N.A.
RX MEDLINE=88209511; PubMed=3365401;
RA Wetsel R.A., Lemons R.S., Lebeau M.M., Barnum S.R., Noack D.,
Tack B.F.:
RT "Molecular analysis of human complement component C5: localization of
the structural gene to chromosome 9.";
Biochemistry 27:1474-1482(1988).
RN 13
RP SEQUENCE OF 412-902 FROM N.A.
RX MEDLINE=85111117; PubMed=2579066;
RA Lundvall J., Wetsel R.A., Kristensen T., Whitehead A.S.,
Woods D.E., Udden R.C., Colten H.R., Tack B.F.:
RT "Isolation and sequence analysis of a cDNA clone encoding the fifth
complement component.";
J. Biol. Chem. 260:2108-2112(1985).
RN 14
RP SEQUENCE OF 678-751.
RX MEDLINE=79005687; PubMed=690134;
RA Fernandez H.N., Hugli T.E.:
RT "Primary structural analysis of the polypeptide portion of human C5a
anaphylatoxin. Polypeptide sequence determination and assignment of
the oligosaccharide attachment site in C5a.";
J. Biol. Chem. 253:6955-6964(1978).
RN 15
RP SEQUENCE OF 678-751 FROM N.A.
RX MEDLINE=91144547; PubMed=1996961;
RA Bonhag J.F., Mollison K.W., Buko A.M., Ashworth J.C., Hill H.R.:
RT "Group B streptococci inactivate complement component C5a by enzymic
cleavage at the C-terminus.";
Biochem. J. 273:635-640(1991).
RN 16
RP STRUCTURE BY NMR OF C5A.
RX MEDLINE=88309754; PubMed=3408713;
RA Zunderweg E.R.P., Mollison K.W., Henkin J., Carter G.W.:
RT "Sequence-specific assignments in the 1H NMR spectrum of the human

RT Inflammatory protein C5a.
 RL Biochemistry 27:3568-3580(1988).
 RN [7]
 RP STRUCTURE BY NMR OF C5A.
 RX MEDLINE-89207527; Pubmed-2784981;
 RA Zunderweg E.R.P., Nellesheim D.G., Mollison K.W., Carter G.W.;
 RT "tertiary structure of human complement component C5a in solution
 from nuclear magnetic resonance data."
 RL Biochemistry 28:172-185(1989).
 RN [8]
 RP STRUCTURE BY NMR OF C5A.
 RX MEDLINE-89274164; Pubmed-2730871;
 RA Zunderweg E.R.P., Fesik S.W.;
 RT "heteronuclear three-dimensional NMR spectroscopy of the inflammatory
 protein C5a."
 RL Biochemistry 28:2387-2391(1989).
 RN [9]
 RP STRUCTURE BY NMR OF C5A.
 RX MEDLINE-97160477; Pubmed-9007977;
 RA Zhang X., Boyar W., Galakatos N., Connella N.C.;
 RT "Solution structure of a unique C5a semi-synthetic antagonist:
 implications in receptor binding."
 RL Protein Sci. 6:65-72(1997).
 RN [10]
 RP STRUCTURE BY NMR OF C5A.
 RX MEDLINE-97332508; Pubmed-9188742;
 RA Zhang X., Boyar W., Toth M.J., Wennogle L., Connella N.C.;
 RT "Structural definition of the C5a C terminus by two-dimensional
 nuclear magnetic resonance spectroscopy."
 RL Proteins 28:261-267(1997).
 CC -1- FUNCTION: ACTIVATION OF C5 BY A C5 CONVERTASE INITIATES THE
 CC SPONTANEOUS ASSEMBLY OF THE LATE COMPLEMENT COMPONENTS, C5-C9,
 CC INTO THE MEMBRANE ATTACK COMPLEX. C5B HAS A TRANSIENT BINDING SITE
 CC FOR C6. THE C5B-C6 COMPLEX IS THE FOUNDATION UPON WHICH THE LYTIC
 CC COMPLEX IS ASSEMBLED.
 CC -1- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C5,
 CC C3 ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT
 CC INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR
 CC PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND
 CC BASOPHILIC LEUKOCYTES. C5A ALSO STIMULATES THE LOCOMOTION OF
 CC POLYMORPHONUCLEAR LEUKOCYTES (CHEMOKINESIS) AND DIRECT THEIR
 CC MIGRATION TOWARD SITES OF INFLAMMATION (CHEMOTAXIS).
 CC -1- SUBUNIT: C5 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 BASIC
 CC RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE
 CC BOND. C5 CONVERTASE ACTIVATES C5 BY CLEAVING THE ALPHA CHAIN,
 CC RELEASING C5A ANAPHYLATOXIN & GENERATING C5B (BETA CHAIN + ALPHA
 CC CHAIN).
 CC -1- SIMILARITY: TO C3, C4 AND ALPHA-2-MACROGLOBULIN.
 CC -1- SIMILARITY: CONTAINS 1 ANAPHYLATOXIN-LIKE DOMAIN.
 CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 855
 CC ONWARD DUE TO THE PRESENCE OF AN ALU REPEAT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M57729; AAAS1925.1;
 DR EMBL: M65134; AAAS1856.1;
 DR PIR: A40075; CSHU.
 DR PIR: S15121; S15121.
 DR PDB: 1KJ5; 15-MAY-97.
 DR PDB: 1CFA; 17-SEP-97.
 DR MIM: 120900;
 DR InterPro: IPR002890; AZM_N.
 DR InterPro: IPR001599; Alpha_2_macroglobulin.
 DR InterPro: IPR000020; Anaphylatoxin.
 DR InterPro: IPR001840; Anaphylatoxin.
 DR InterPro: IPR001134; Netrin_C.
 DR Pfam: PF00207; AZM; 1.

DR Pfam: PF01835; AZM_N; 1.
 DR Pfam: PF01831; ANATO; 1.
 DR Pfam: PF01759; NTR; 1.
 DR PRINTS: PRO00004; ANAPHYLATOXN.
 DR SMART: SM00104; ANATO; 1.
 DR PROSITE: PS00477; ALPHA_2_MACROGLOBULIN; FALSE_NEG.
 DR PROSITE: PS01177; ANAPHYLATOXIN_1; 1.
 DR PROSITE: PS01178; ANAPHYLATOXIN_2; 1.
 KM Complement pathway: Complement alternate pathway: Glycoprotein;
 KM plasma: Membrane attack complex: Cytolysis: Inflammatory response;
 KM signal: Polymorphism: 3D-structure.
 FT SIGNAL 1 18
 FT CHAIN 19 673
 FT PROPEP 674 677
 FT CHAIN 678 1676
 FT PEPTIDE 678 751
 FT CHAIN 752 1676
 FT DOMAIN 698 732
 FT DISULFID 699 731
 FT DISULFID 711 732
 FT CARBOHYD 741 741
 FT CARBOHYD 911 911
 FT CARBOHYD 1115 1115
 FT CARBOHYD 1630 1630
 FT VARIANT 518 518
 FT SEQUENCE 1676 AA; 188331 MW; 87DCAA65FF977D19 CRC64;
 SQ
 Query Match 100.0%; Score 52; DB 1; Length 1676;
 Best Local Similarity 100.0%; Pred. No. 0.012;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Caps 0;
 Oy 1 ISHKDMLGR 10
 Db 742 ISHKDMLGR 751
 RESULT 2
 OTCA_PSESH STANDARD: PRT: 306 AA.
 AD Q02047;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE ORNITHINE CARBAMOYLTRANSFERASE, PHASELOTOXIN-SENSITIVE, ANABOLIC
 DE (EC 2.1.3.3) (OTCASE).
 GN ARCF.
 OS Pseudomonas syringae (pv. phaseolicola).
 OC Bacteria: Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=319;
 RN [1]
 RN SEQUENCE FROM N.A.
 RA Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.
 RX MEDLINE-92394893; Pubmed-1522066;
 RA Hatziloukas E., Panopoulos N.J.;
 RT "Origin, structure, and regulation of argK, encoding the
 RT phaseolotoxin-resistant ornithine carbamoyltransferase in
 RT Pseudomonas syringae pv. phaseolicola, and functional expression of
 RT argK in transgenic tobacco."
 RL J. Bacteriol. 174:5895-5909(1992).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN-MEXICAN ISOLATE TEXCOCO;
 RA Kosquenda-Cano G.;
 CC -1- CATALYTIC ACTIVITY: CARBAMOYL PHOSPHATE + ORNITHINE = CITRULLINE
 CC + ORTHOPHOSPHATE.
 CC -1- PATHWAY: ARGININE BIOSYNTHESIS.
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.

```

-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: M93182; AAA25909.1; -
DR  EMBL: X76945; CA54264.1; -
DR  PIR: S41291; S41291.
DR  HSSP: O51742; 1A1S.
DR  InterPro: IPR002029; Carboxyltransferase.
DR  Pfam: PF00185; OTCace; 1.
DR  Pfam: PF02729; OTCace_N; 1.
DR  PRINTS: PR00100; AOTCASE.
DR  PRINTS: PR00102; OTCASE.
DR  PROSITE: PS00097; CARBAMOYLTRANSFERASE; 1.
KW  Arginine biosynthesis; Transferase.
SQ  SEQUENCE 306 AA: 34151 MW: D37B503563D53F50 CRC64:

```

```

Query Match          78.8%; Score 41; DB 1; Length 306;
Best Local Similarity 70.0%; Pred. No. 0.37;
Matches 7: Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 1 ISHKMOLGR 10
    11:11111
Db 74 LSHRDTOLGR 83

```

```

RESULT 3
COSY_BOVIN STANDARD: PRT: 74 AA.
ID COSY_BOVIN STANDARD: PRT: 74 AA.
PI2082;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE COMPLEMENT C5A ANAPHYLATOXIN.
CN C5.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID:9913;
RN [1]
RP SEQUENCE.
RX MEDLINE=86136134; PubMed=3081348;
RA Gennaro R., Simonic T., Negri A., Mottola C., Secchi C., Ronchi S.,
RA Romeo D.;
RT "C5a fragment of bovine complement. Purification, bioassays,
RT amino-acid sequence and other structural studies.";
RL Eur. J. Biochem. 155:77-86(1986).
RN [2]
RP STRUCTURE BY NMR.
RX MEDLINE=89005703; PubMed=3262536;
RA Zardock J., Gennaro R., Romeo D., Clore G.M., Gronenborn A.M.;
RT "A proton nuclear magnetic resonance study of the conformation of
RT bovine anaphylatoxin C5a in solution.";
RL FEBS Lett. 238:289-294(1988).
CC -1- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C5,
CC C5 ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT
CC INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR
CC PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND
CC BASOPHILIC LEUKOCYTES. C5A ALSO STIMULATES THE LOCOMOTION OF
CC POLYMORPHONUCLEAR LEUKOCYTES (CHEMOKINESIS) AND DIRECT THEIR
CC MIGRATION TOWARD SITES OF INFLAMMATION (CHEMOTAXIS).
CC -1- SIMILARITY: CONTAINS 1 ANAPHYLATOXIN-LIKE DOMAIN.
DR PIR: A25408; A25408.
DR HSSP: P01032; 1C5A.
DR InterPro: IPR001599; Alpha_2_macrogllobln.
DR InterPro: IPR000020; Anaphylatoxin.

```

```

DR InterPro: IPR001840; Anaphylatoxn.
DR Pfam: PF01821; ANATO; 1.
DR PRINTS: PR00004; ANAPHYLATOXN.
DR SMART: SM00104; ANATO; 1.
DR PROSITE: PS00477; ALPHA_2_MACROGLOBULIN_PARTIAL.
DR PROSITE: PS01177; ANAPHYLATOXIN_1; 1.
DR PROSITE: PS01178; ANAPHYLATOXIN_2; 1.
KW Complement pathway; Complement alternate pathway; Plasma;
KW Inflammatory response.
FT DOMAIN 21 55 ANAPHYLATOXIN-LIKE.
FT DISULFD 21 47 BY SIMILARITY.
FT FT DISULFD 22 54 BY SIMILARITY.
FT FT DISULFD 34 55 BY SIMILARITY.
SQ SEQUENCE 74 AA: 8517 MW: C09DF742D12D70F6 CRC64:

```

```

Query Match          75.0%; Score 39; DB 1; Length 74;
Best Local Similarity 87.5%; Pred. No. 0.2;
Matches 7: Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 3 HKDMOLGR 10
    11:11111
Db 67 HKNMOLGR 74

```

```

RESULT 4
COSY_PIG STANDARD: PRT: 74 AA.
ID COSY_PIG STANDARD: PRT: 74 AA.
PI01032;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE COMPLEMENT C5A ANAPHYLATOXIN.
CN C5.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE.
RX MEDLINE=80182137; PubMed=7372604;
RA Gerard C., Hugli T.E.;
RT "Amino acid sequence of the anaphylatoxin from the fifth component of
RT porcine complement";
RL J. Biol. Chem. 255:4710-4715(1980).
RN [2]
RP ACTIVE REGION.
RX MEDLINE=81199549; PubMed=6940191;
RA Gerard C., Hugli T.E.;
RT "Identification of classical anaphylatoxin as the des-Arg form of the
RT C5a molecule: evidence of a modulator role for the oligosaccharide
RT unit in human des-Arg74-C5a.";
RL Proc. Natl. Acad. Sci. U.S.A. 78:1833-1837(1981).
RN [3]
RP STRUCTURE BY NMR.
RX MEDLINE=90248365; PubMed=2337573;
RA Williamson M.P., Madison V.S.;
RT "Three-dimensional structure of porcine C5adesArg from 1H nuclear
RT magnetic resonance data.";
RL Biochemistry 29:2895-2905(1990).
CC -1- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C5,
CC C5 ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT
CC INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR
CC PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND
CC BASOPHILIC LEUKOCYTES. C5A ALSO STIMULATES THE LOCOMOTION OF
CC POLYMORPHONUCLEAR LEUKOCYTES (CHEMOKINESIS) AND DIRECT THEIR
CC MIGRATION TOWARD SITES OF INFLAMMATION (CHEMOTAXIS).
CC -1- SIMILARITY: CONTAINS 1 ANAPHYLATOXIN-LIKE DOMAIN.
DR PIR: A01268; C5PGAT.
DR PDB: 1C5A; 15-OCT-91.
DR InterPro: IPR001599; Alpha_2_macrogllobln.
DR InterPro: IPR000020; Anaphylatoxin.
DR InterPro: IPR001840; Anaphylatoxn.

```

DR Pfam: PF01821; ANATO; 1.
 DR PRINTS: PR00004; ANAPHYLATOXN.
 DR SMART: SM00104; ANATO; 1.
 DR PROSITE: PS00477; ALPHA_2_MACROGLOBULIN; PARTIAL.
 DR PROSITE: PS01178; ANAPHYLATOXIN_1; 1.
 DR PROSITE: PS01178; ANAPHYLATOXIN_2; 1.
 KW Complement pathway: Complement alternate pathway; Plasma;
 KW Inflammatory response; 3D-structure.
 FT DOMAIN 21 55 ANAPHYLATOXIN-LIKE.
 FT DISULFID 21 54
 FT DISULFID 22 54
 FT DISULFID 34 55
 FT SITE 72 74
 FT HELIX 2 11
 FT TURN 13 14
 FT HELIX 16 26
 FT HELIX 34 40
 FT HELIX 45 62
 FT TURN 63 64
 SO SEQUENCE 74 AA: 8609 MW: 11AAFE2E94A026EB3 CRC64;
 Query Match 75.0%; Score 39; DB 1; Length 74;
 Best Local Similarity 77.8%; Pred. No. 0.2;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 Oy 2 SHKDMQLGR 10
 1111111111
 Db 66 SHKNIOQLGR 74
 RESULT 5
 ID OTCA_BACSU STANDARD: PRT: 319 AA.
 AC P18186;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ORNITHINE CARBAMOYLTRANSFERASE, ANABOLIC (EC 2.1.3.3) (OTCACE).
 GN ARGF.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=168 / EMC50;
 RA MEDLINE=90356402; PubMed=2117745;
 RA Mountain A., Smith M.C.M., Baumberg S.;
 RT "Nucleotide sequence of the Bacillus subtilis argf gene encoding
 RT ornithine carbamoyltransferase.";
 RL Nucleic Acids Res. 18:4594-4594(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=94297722; PubMed=8025667;
 RA O'Reilly M., Devine K.M.;
 RT "Sequence and analysis of the citrulline biosynthetic operon argC-F
 RT from Bacillus subtilis.";
 RL Microbiology 140:1023-1025(1994).
 RN [3]
 RP SEQUENCE OF 82-319 FROM N.A.
 RC STRAIN=CU741;
 RX MEDLINE=97474245; PubMed=9335269;
 RA Ogura M., Ohshiro Y., Hiroo S., Tanaka T.;
 RT "A new Bacillus subtilis gene, med, encodes a positive regulator of
 RT comK.";
 RL J. Bacteriol. 179:6244-6253(1997).
 CC -1- CATALYTIC ACTIVITY: CARBAMOYL PHOSPHATE + ORNITHINE - CITRULLINE
 CC + ORTHOPHOSPHATE.
 CC -1- PATHWAY: ARGININE BIOSYNTHESIS.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.

CC -1- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X53360; CAA37444.1; -
 DR EMBL: Z26919; CAA81542.1; -
 DR EMBL: Z29109; CAB12966.1; -
 DR EMBL: Z29110; CAB12982.1; -
 DR EMBL: D86376; BAA22924.1; -
 DR PIR: S11000; OMBS.
 DR PIR: S38434; S38434.
 DR HSSP: O51742; 1A1S.
 DR Subtilist; BG10197; argf.
 DR InterPro: IPR002029; Carbmyltransf_asor.
 DR Pfam: PF00185; OTCace; 1.
 DR Pfam: PF02729; OTCace.N; 1.
 DR PRINTS: PR00100; AOTCACE.
 DR PRINTS: PR00102; OTCACE.
 DR PROSITE: PS00097; CARBAMOYLTRANSFERASE; 1.
 KW Transferase; Arginine biosynthesis; Complete proteome.
 SO SEQUENCE 319 AA: 34663 MW: E3EE18A5A9BD89DE CRC64;
 Query Match 75.0%; Score 39; DB 1; Length 319;
 Best Local Similarity 70.0%; Pred. No. 1;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 ISHKDMQLGR 10
 1111111111
 Db 78 LSQKDQLGR 87
 RESULT 6
 ID RPC2_YEAST STANDARD: PRT: 1149 AA.
 AC P22276; Q12696;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE DNA-DIRECTED RNA POLYMERASE III 130 KDA POLYPEPTIDE (EC 2.7.7.6)
 DE (C128) (RNA POLYMERASE III SUBUNIT 2).
 GN RPC2 OR RPC128 OR RET1 OR YOR207C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91170230; PubMed=2005101;
 RA James P., Whelen S., Hall B.D.;
 RT "The RET1 gene of yeast encodes the second-largest subunit of RNA
 RT polymerase III. Structural analysis of the wild-type and ret-1
 RT mutant alleles.";
 RL J. Biol. Chem. 266:5616-5624(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Hughes B., Pohl T.M.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 CC SUBSTRATES.
 CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +
 CC RNA(N).
 CC -1- SUBUNIT: RNA POLYMERASE III CONSISTS OF ABOUT 15 DIFFERENT
 CC SUBUNITS. THIS SUBUNIT IS THE SECOND LARGEST COMPONENT OF RNA
 CC POLYMERASE III.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.


```

CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC PRECURSOR, POLYMERASE II FOR THE mRNA PRECURSOR, AND POLYMERASE
CC III FOR 5S AND TRNA GENES.
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified, and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M38723; AAB59324.1; -.
DR EMBL: 275115; CA99422.1; -.
DR PIR: S14169; S14169.
DR SCD: S0005733; RET1.
DR InterPro: IPR001572; RNA_POL_B.
DR Pfam: PF00562; RNA_POL_B.1.
DR PROSITE: PS01166; RNA_POL_BETA.1.
DR Transferrase: DNA-directed RNA polymerase; Transcription; Zinc;
DR Zinc-finger: Nuclear protein.
DR ZN_FING 1095 1110 C4-TYPE (POTENTIAL).
FT CONFLICT 213 213 K -> E (IN REF. 1).
FT SEQUENCE 1149 AA: 129456 MW: 6AC52354F34CF090 CRC64:

Query Match 73.1%; Score 38; DB 1; Length 1149;
Best Local Similarity 60.0%; Pred. No. 6.6;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Oy 1 ISKDMOLCR 10
Db 142 IMKQVEIGR 151

RESULT 7
COSA_RAT STANDARD: PRT; 76 AA.
AC P08650;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE COMPLEMENT C5A ANAPHYLATOXIN.
GN C5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE.
RA Cul L.-X., Ferreri K., Hugli T.E.;
RT *Characterization of rat C5a, a uniquely active spasmogen.*;
RL Complement 2:18-19(1985).
CC -1- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C5,
CC C5 ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT
CC INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR
CC PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND
CC BASOPHILIC LEUKOCYTES. C5A ALSO STIMULATES THE LOCOMOTION OF
CC POLYMORPHONUCLEAR LEUKOCYTES (CHEMOKINESIS) AND DIRECT THEIR
CC MIGRATION TOWARD SITES OF INFLAMMATION (CHEMOTAXIS).
CC -1- SIMILARITY: CONTAINS 1 ANAPHYLATOXIN-LIKE DOMAIN.
CC HSSP: P01031; 1KJ5.
DR InterPro: IPR001599; Alpha_2_macroglbln.
DR InterPro: IPR000020; Anaphylatoxin.
DR InterPro: IPR001840; Anaphylatoxin.
DR Pfam: PF01821; ANATO.1.
DR PRINTS: PM00004; ANAPHYLATOXN.
DR SMART: SM00104; ANATO.1.
DR PROSITE: PS00477; ALPHA_2_MACROGLOBULIN; PARTIAL.
DR PROSITE: PS01177; ANAPHYLATOXIN_1; 1.
DR PROSITE: PS01178; ANAPHYLATOXIN_2; 1.

```

```

KW Complement pathway; Complement alternate pathway; Glycoprotein;
KM Plasma; Inflammatory response.
FT DOMAIN 24 58 ANAPHYLATOXIN-LIKE.
FT DISULFID 24 50 BY SIMILARITY.
FT DISULFID 25 57 BY SIMILARITY.
FT DISULFID 37 58 BY SIMILARITY.
FT CARBOHYD 66 66 N-LINKED (GLCNAC. . .).
SO SEQUENCE 76 AA: 8869 MW: 2EC15183AE54769 CRC64:

Query Match 65.4%; Score 34; DB 1; Length 76;
Best Local Similarity 77.8%; Pred. No. 2.3;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 2 SHKDMOLCR 10
Db 68 SHKGMILCR 76

RESULT 8
CYPD_YEAST STANDARD: PRT; 225 AA.
ID CYPD_YEAST
AC P35176;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE PERTIDYL-PROLYL CIS-TRANS ISOMERASE D PRECURSOR (EC 5.2.1.8) (PIPIASE)
DE (ROTAMASE) (CYCLOPHILIN D).
GN CPR5 OR CYP5 OR CYPD OR YOR304C OR D9740.14.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
CX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93389734; PubMed=8377189;
RA Frigerio G., Pelham H.R.B.;
RT "A Saccharomyces cerevisiae cyclophilin resident in the endoplasmic
RT reticulum.";
RL J. Mol. Biol. 233:183-188(1993).
FN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favella A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jor M.,
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Madis E., Meneses S., Miller N., Nnan M., Pauley A., Peluso D.,
RA Ritken L., Riles L., Tatch A., Trevasakis E., Vignati D.,
RA Wilcox L., Wohlman P., Vaudin M., Wilson R., Waterston R.;
RL Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: PIPIASES ACCELERATE THE FOLDING OF PROTEINS.
CC -1- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC
CC PEPTIDE BONDS IN OLIGOPEPTIDES.
CC -1- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN.
CC -1- SIMILARITY: BELONGS TO THE CYCLOPHILIN-TYPE PIPIASE FAMILY.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X73142; CA51658.1; -.
DR EMBL: U28374; AAB64740.1; -.
DR PIR: S38324; S38324.
DR HSSP: P23284; ICYN.
DR SWISS-2DPAGE: P35176; YEAST.
DR YEPD: 8236; -.
DR SGD: S0002712; CYP5.
DR InterPro: IPR002130; CSA_Piase.

```

```

DR InterPro: IPR000886: ER-target.
DR Pfam: PF00160: PRO_Isomerase: 1.
DR PRINTS: PR00153: CSAPISMASE.
DR PROSITE: PS00014: ER_TARGET: 1.
DR PROSITE: PS00170: CSA_PPIASE.1: 1.
DR PROSITE: PS50072: CSA_PPIASE.2: 1.
DR Cyclosporin: Isomerase: Rotamase: Signal: Endoplasmic reticulum:
KW Multigene family.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 225 PEPTIDYL-PROLYL CIS-TRANS ISOMERASE D.
FT SITE 222- 225 PREVENT SECRETION FROM ER.
FT CARBOHYD: 139 139 N-LINKED (GLCNAC...) (POTENTIAL).
SO SEQUENCE: 225 AA: 25327 MW: F4861424C8443B58 CRC64;

Query Match
Best Local Similarity 65.4%: Score 34: DB 1: Length 225;
Matches 6: Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 ISHKMOLGR 10
Db 40 INHCKOIGR 49

RESULT 9
BPHG_BURCE
ID BPHG_BURCE STANDARD: PRT: 408 AA.
AC P37337:
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE BI-PHENYL DIOXYGENASE SYSTEM FERREDOXIN--NAD(+) REDUCTASE COMPONENT
DE (EC 1.18.1.3).
GN BPHG.
OS Burkholderia cepacia (Pseudomonas cepacia).
OC Bacteria: Proteobacteria: beta subdivision: Burkholderia group:
OC Burkholderia.
OC NCBI_TaxID=292;
OX [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=92234948: Pubmed=1569021:
RX STRAIN=LB400:
RA Erickson B.D., Mondello F.J.:
RT "Nucleotide sequencing and transcriptional mapping of the genes
RT encoding bi-phenyl dioxygenase, a multicomponent
RT polychlorinated-bi-phenyl-degrading enzyme in Pseudomonas strain
RT LB400."
RT J. Bacteriol. 174:2903-2912(1992).
RN [2]
RN REVISIONS.
RA Erickson B.D., Mondello F.J.:
RA Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PART OF THE ELECTRON TRANSFER COMPONENT OF BI-PHENYL
CC DIOXYGENASE. TRANSFERS ELECTRONS FROM FERREDOXIN (BPHF) TO
CC NADH.
CC -1- CATALYTIC ACTIVITY: REDUCED FERREDOXIN + NAD(+) -> OXIDIZED
CC FERREDOXIN + NADH.
CC -1- COFACTOR: FAD.
CC -1- PATHWAY: BI-PHENYL-POLYCHLORINATED BI-PHENYL DEGRADATION PATHWAY.
CC -1- SUBUNIT: THIS DIOXYGENASE SYSTEM CONSISTS OF FOUR PROTEINS: THE
CC TWO SUBUNITS OF THE HYDROXYLASE COMPONENT (BPHA AND BPHF), A
CC FERREDOXIN (BPHF) AND A FERREDOXIN REDUCTASE (BPHG).
CC -1- SIMILARITY: TO OTHER BACTERIAL RING-HYDROXYLATING DIOXYGENASE
CC FERREDOXIN REDUCTASE COMPONENTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR EMBL: M86348: AAB63429.1: -.
DR PIR: F41858: F41858.
DR InterPro: IPR001327: FAD_Pyr-redox.
DR InterPro: IPR00205: NAD_binding.
DR Pfam: PF00070: Pyr_redox: 1.
DR PRINTS: PR00368: FADPNR.
KW Aromatic hydrocarbons catabolism: Flavoprotein: FAD: NAD:
KW Oxidoreductase.
FT NP_BIND 4 35 FAD (ADP PART) (POTENTIAL).
FT NP_BIND 145 173 NAD (ADP PART) (POTENTIAL).
SO SEQUENCE 408 AA: 42953 MW: 8A52B0168667A9 CRC64;

Query Match
Best Local Similarity 66.7%: Score 34: DB 1: Length 408;
Matches 6: Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 SHKMOLGR 10
Db 70 AHVDVQLGR 78

RESULT 10
YMXI_CAEEL
ID YMXI_CAEEL STANDARD: PRT: 547 AA.
AC P34509:
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE SERINE/THREONINE-PROTEIN KINASE K06H7.1 IN CHROMOSOME III
DE (EC 2.7.1.1).
GN K06H7.1.
OS Caenorhabditis elegans.
OC Eukaryota: Metazoa: Nematoda: Chromadorea: Rhabditida: Rhabditioidea:
OC Rhabditidae: Pelodierinae: Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=BRISTOL N2:
RX MEDLINE=94150718: Pubmed=7906398:
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Letellier P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Koopra A., Saunders D., Shonkeen R.,
RA Sims M., Smailton N., Smith A., Smith M., Sonhammer B., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Watson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
PA Mohlman P.:
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RT Nature 368:32-38(1994).
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L15314: AAA28084.1: -.
CC PIR: S44841: S44841.
CC HSSP: Q63450: 1A06.
CC WormPep: K06H7.1: CE00252.
DR InterPro: IPR00719: Euk_pkinase.
DR InterPro: IPR002290: Ser_thr_kin_actslve.
DR Pfam: PF00069: pkinase: 1.
DR SMART: SM00220: S_TKc.1.
DR PROSITE: PS00107: PROTEIN_KINASE_ATP; 1.

```

DR PROSITE; PS00108; PROTEIN_KINASE-ST; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE-DOM; 1.
 KW Hypothetical protein; Transferase; Serine/threonine-protein kinase:
 KM ATP-binding. 267 526 PROTEIN KINASE.
 FT DOMAIN 273 281 ATP (BY SIMILARITY).
 FT NP_BIND 273 281 ATP (BY SIMILARITY).
 FT BINDING 296 296 ATP (BY SIMILARITY).
 FT ACT_SITE 390 390 BY SIMILARITY.
 SQ SEQUENCE 547 AA: 63490 MW: 10CD28C2FEAC63101 CRC64:

Query Match 65.4%; Score 34; DB 1; Length 547;
 Best Local Similarity 66.7%; Pred. No. 20;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 ISHKDMOLG 9
 DB 386 IIRHDMKLG 394

RESULT 11
 ID YFYA_SCHPO STANDARD: PRT: 591 AA.
 AC 09UT18:
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PUTATIVE AMINO-ACID PERMEASE C9.10.
 GN SPAC9.10.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;
 CC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Weiler H., Duesterhoeft A., Lyne M.H., Rajandream M.A., Barrell B.G.;
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AL121764; CAB57428.1; -
 DR InterPro: IPR002293; AA_rel-permease_1;
 DR InterPro: IPR002027; Amino_acid-permease.
 DR Pfam: PF00324; aa-permeases; 1.
 DR PROSITE; PS00218; AMINO-ACID-PERMEASE; FALSE-NEG.
 KW Hypothetical protein; Transport; Amino-acid transport.
 KM Transmembrane.
 FT TRANSMEM 98 118 POTENTIAL.
 FT TRANSMEM 342 362 POTENTIAL.
 FT TRANSMEM 397 417 POTENTIAL.
 FT TRANSMEM 450 470 POTENTIAL.
 FT TRANSMEM 545 565 POTENTIAL.
 SQ SEQUENCE 591 AA: 63069 MW: FFCDD707AC7849D4 CRC64:

Query Match 65.4%; Score 34; DB 1; Length 591;
 Best Local Similarity 60.0%; Pred. No. 22;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 ISHKDMOLGR 10
 DB 6 ISHODPELCO 15

RESULT 12
 ID PURL_BACHD STANDARD: PRT: 743 AA.
 AC 09KE57:
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PHOSPHORIBOSYLFORMYLGLYCINAMIDE SYNTHASE II (EC 6.3.5.3) (FGAM
 DE SYNTHASE II)
 GN PURL OR BH0629.
 OS Bacillus halodurans.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=8565;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.,
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 RT halodurans and genomic sequence comparison with Bacillus subtilis.";
 RL Nucleic Acids Res. 28:4317-4331(2000).
 CC -1- CATALYTIC ACTIVITY: ATP + 5'-PHOSPHORIBOSYLFORMYLGLYCINAMIDE +
 CC L-GLUTAMINE + H(2)O = ADP + ORTHOPHOSPHATE + 5'-PHOSPHORIBOSYL-
 CC FORMYLGLYCINAMIDE + L-GLUTAMATE.
 CC -1- PATHWAY: DE NOVO PURINE BIOSYNTHESIS; FOURTH STEP.
 CC -1- SUBUNIT: HETERODIMER OF TWO SUBUNITS, PURO AND PURL.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE FGAMS FAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AP001509; BAB04348.1; -
 DR InterPro: IPR000728; ATRS-related.
 DR InterPro: IPR002086; Aldehyde-dehydr.
 DR Pfam: PF00586; ATRS; 2.
 KW Purine biosynthesis; Ligase; ATP-binding; Complete proteome.
 FT NP_BIND 111 122 ATP (POTENTIAL).
 SQ SEQUENCE 743 AA: 80008 MW: 005FC0855D20D84D CRC64:

Query Match 65.4%; Score 34; DB 1; Length 743;
 Best Local Similarity 60.0%; Pred. No. 28;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 ISHKDMOLGR 10
 DB 194 IDHKDOKGO 203

RESULT 13
 ID MRP2_RAT STANDARD: PRT: 1541 AA.
 AC 063120; 063145;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE CANALICULAR MULTISPECIFIC ANION TRANSPORTER 1 (MULTIDRUG
 DE RESISTANCE-ASSOCIATED PROTEIN 2) (CANALICULAR MULTIDRUG RESISTANCE
 DE PROTEIN).
 GN ABCC2 OR CMOAT OR MRP2 OR CMRP.
 GN Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

CC NCBI_TaxID=10116;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN-WISTAR: TISSUE=Liver;
 RX MEDLINE=96180672; PubMed=8599091;
 RA Paulusma C.C., Bosma P.J., Zaman G.J.R., Bakker C.T.M., Otter M.,
 RA Schetter G.L., Schepers R.J., Borst P., Oude Elferink R.P.J.;
 RT "Congenital jaundice in rats with a mutation in a multidrug
 RT resistance-associated protein gene.";
 RL Science 271:1126-1128(1996).
 RN [2]
 RE SEQUENCE FROM N.A.
 RC STRAIN-WISTAR: TISSUE=Liver;
 RX MEDLINE=96279006; PubMed=8662992;
 RA Buechler D., Koenig J., Brom M., Kartenbeck J., Spring H., Horie T.,
 RA Keppler D.;
 RT "cDNA cloning of the hepatocyte canalicular isoform of the multidrug
 RT resistance protein, cMRP, reveals a novel conjugate export pump
 RT deficient in hyperbilirubinemic mutant rats.";
 RL J. Biol. Chem. 271:15091-15098(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY: TISSUE=Liver;
 RA Ito K., Suzuki H., Hirohashi T., Kume K., Shimizu T., Sugiyama Y.;
 RT "Expression of the putative ATP-binding cassette region, homologous to
 RT that in multidrug resistance associated protein (MRP), is hereditarily
 RT defective in Elsal hyperbilirubinemic rats (EHBR).";
 RL Hepatol. Commun. 292:292-299(1996).
 CC -1- FUNCTION: MEDIATES HEPATOBLILIARY EXCRETION OF NUMEROUS ORGANIC
 CC ANIONS.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: MAINLY EXPRESSED IN THE LIVER.
 CC -1- DISEASE: DEFECTS IN MRP2 ARE A CAUSE OF HEREDITARY CONJUGATED
 CC HYPERBILIRUBINEMIA (EHBR).
 CC -1- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
 CC (ABC TRANSPORTERS). MRP SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: LA9379; AAC42087.1; -;
 DR EMBL: X96393; CA65257.1; -;
 DR EMBL: D86086; BAA13016.1; -;
 DR HSSP: P13569; INBD.
 DR InterPro: IPR003593; AAA.
 DR InterPro: IPR003439; ABC_Transport.
 DR InterPro: IPR001140; ABC_Transport_tmem.
 DR InterPro: IPR001687; ATP_CTP_A.
 DR Pfam: PF00064; ABC_membrane; 2.
 DR Pfam: PF00005; ABC_tran; 2.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 KM ATP-binding: Glycoprotein; Transmembrane; Transport; Repeat.
 FT DOMAIN 1 26
 FT TRANSEM 1 26
 FT DOMAIN 27 47
 FT TRANSEM 27 47
 FT DOMAIN 48 67
 FT TRANSEM 48 67
 FT DOMAIN 68 88
 FT TRANSEM 68 88
 FT DOMAIN 89 92
 FT TRANSEM 89 92
 FT DOMAIN 93 113
 FT TRANSEM 93 113
 FT DOMAIN 114 125
 FT TRANSEM 114 125
 FT DOMAIN 126 146
 FT TRANSEM 126 146
 FT DOMAIN 147 164
 FT TRANSEM 147 164
 FT DOMAIN 165 185
 FT TRANSEM 165 185
 FT DOMAIN 186 309
 FT TRANSEM 186 309
 FT DOMAIN 310 330
 FT TRANSEM 310 330
 FT DOMAIN 331 356
 FT TRANSEM 331 356
 FT DOMAIN 357 377
 FT TRANSEM 357 377
 FT DOMAIN 378 433
 FT TRANSEM 378 433

FT TRANSEM 434 454
 FT DOMAIN 455 457
 FT TRANSEM 458 478
 FT DOMAIN 479 540
 FT TRANSEM 541 561
 FT DOMAIN 562 583
 FT TRANSEM 584 604
 FT DOMAIN 605 667
 FT TRANSEM 668 688
 FT DOMAIN 689 1029
 FT TRANSEM 1030 1050
 FT DOMAIN 1051 1093
 FT TRANSEM 1094 1114
 FT DOMAIN 1115 1115
 FT TRANSEM 1116 1136
 FT DOMAIN 1137 1207
 FT TRANSEM 1208 1228
 FT DOMAIN 1229 1230
 FT TRANSEM 1231 1251
 FT DOMAIN 1252 1541
 FT NP_BIND 667 674
 FT NP_BIND 674 674
 FT CARBOHYD 1330 1337
 FT CARBOHYD 6 6
 FT CARBOHYD 1007 1007
 FT CARBOHYD 1010 1010
 FT CARBOHYD 1011 1011
 FT CONFLICT 420 420
 SQ SEQUENCE 1541 AA; 173383 MW; D5FB55571BFDB39 CRC64;
 M -> V (IN REF. 3).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 M -> V (IN REF. 3).
 D5FB55571BFDB39 CRC64;
 Query Match 65.4%; Score 34; DB 1; Length 1541;
 Best Local Similarity 62.5%; Pred. No. 62;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 2 SHKMOIC 9
 DB 1014 SHKMOIC 1021
 RESULT 14
 ID COS_MOUSE STANDARD: PRT: 1680 AA.
 DC COS_MOUSE
 DC P06684;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE COMPLEMENT C5 PRECURSOR (HEMOLYTIC COMPLEMENT) [CONTRAINS: C5A
 DE ANAPHYLATOXIN].
 GN C5 OR HC.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=96279006; PubMed=8662992;
 RA Wetsel R.A., Ogata R.T., Tack B.F.;
 RT "Deficiency of the murine fifth complement component (C5). A 2-base
 RT pair gene deletion in a 5'-exon";
 RL J. Biol. Chem. 265:2435-2440(1990).
 RN [2]
 RE SEQUENCE OF 41-1680 FROM N.A.
 RC MEDLINE=87185363; PubMed=2436653;
 RX Wetsel R.A., Ogata R.T., Tack B.F.;
 RT "Primary structure of the fifth component of murine complement.";
 RL Biochemistry 26:737-743(1987).
 CC -1- FUNCTION: ACTIVATION OF C5 BY A C5 CONVERTASE INITIATES THE
 CC SPONTANEOUS ASSEMBLY OF THE LATE COMPLEMENT COMPONENTS, C5-C9,
 CC INTO THE MEMBRANE ATTACK COMPLEX. C5B HAS A TRANSIENT BINDING SITE
 CC FOR C6. THE C5B-C6 COMPLEX IS THE FOUNDATION UPON WHICH THE LYTIC
 CC COMPLEX IS ASSEMBLED.
 CC -1- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C5,
 CC C5 ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT

INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND BASOPHILIC LEUKOCYTES. C5A ALSO STIMULATES THE LOCOMOTION OF POLYMORPHONUCLEAR LEUKOCYTES (CHEMOKINESIS) AND DIRECT THEIR MIGRATION TOWARD SITES OF INFLAMMATION (CHEMOTAXIS).

-1- SUBUNIT: C5 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 BASIC RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE BOND. C5 CONVERTASE ACTIVATES C5 BY CLEAVING THE ALPHA CHAIN, RELEASING C5A ANAPHYLATOXIN & GENERATING C5B (BETA CHAIN + ALPHA CHAIN).

-1- SIMILARITY: TO C3, C4 AND ALPHA-2-MACROGLOBULIN.

-1- SIMILARITY: CONTAINS 1 ANAPHYLATOXIN-LIKE DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: M35525; AAA37349.1; -
 CC EMBL: M35526; AAA37348.1; -
 CC PIR: A27538; A27538.
 CC PIR: A35530; A35530.
 CC HSSP: P01031; IKSJ.
 CC MCD: MGI:96031; HC.
 DR InterPro: IPR002890; A2M_N.
 DR InterPro: IPR001599; Alpha_2_macroglobin.
 DR InterPro: IPR000020; Anaphylatoxin.
 DR InterPro: IPR001840; Anaphylatoxin.
 DR InterPro: IPR001134; Netrin_C.
 DR Pfam: PF00207; A2M_L.
 DR Pfam: PF01835; A2M_N; 1.
 DR Pfam: PF01821; ANATO; 1.
 DR Pfam: PF01759; NTR; 1.
 DR PRINTS: PR00004; ANAPHYLATOXN.
 DR SMART: SM00104; ANAPHYLATOXN.
 DR PROSITE: PS00477; ALPHA_2_MACROGLOBULIN; FALSE_NEG.
 DR PROSITE: PS01177; ANAPHYLATOXIN_1; 1.
 DR PROSITE: PS01178; ANAPHYLATOXIN_2; 1.
 KW Complement pathway: Complement alternate pathway; Glycoprotein;
 KW Plasma: Membrane attack complex; Cytolysis; Inflammatory response;
 KW Signal.

FT SIGNAL 1 18
 FT CHAIN 19 1680
 FT CHAIN 19 674
 FT PROPEP 675 678
 FT CHAIN 679 1680
 FT PEPTIDE 679 755
 FT CHAIN 756 1680
 FT DOMAIN 702 736
 FT DISULFID 702 728
 FT DISULFID 703 735
 FT DISULFID 715 736
 FT CARBOHYD 427 427
 FT CARBOHYD 915 915
 FT CARBOHYD 1119 1119
 FT CARBOHYD 1633 1633
 FT VARIANT 216 216
 FT VARIANT 217 1680
 FT SEQUENCE 1680 AA: 188877 MW: 81EB5A16FAC7D95C CRC64:

Query Match 63.5%; Score 33; DB 1; Length 1680;
 Best Local Similarity 75.0%; Pred. No. 1,1e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 HKDMQLGR 10
 11:1111
 DB 748 HKPVQLGR 755

RESULT 15
 YC11_METJA
 ID YC11_METJA STANDARD; PRT; 116 AA.
 AC 056608;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL PROTEIN MJ1211.
 GN MJ1211.
 OS Methanococcus jannaschii.
 CC Archaea: Euryarchaeota; Methanococcales; Methanococcaceae;
 CC Methanococcus.
 CC NCBI_TaxID=2190;
 CX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN:JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodok A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.V., Nguyen D.,
 RA Oltnerback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT *Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii.*;
 RL Science 273:1058-1073(1996).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: U67562; AAB99222.1; -
 DR TIGR: MJ1211; -
 KW Hypothetical protein; Complete proteome.
 KW SEQUENCE 116 AA: 13422 MW: 0EA61BCFCEAA7C CRC64;

Query Match 61.5%; Score 32; DB 1; Length 116;
 Best Local Similarity 60.0%; Pred. No. 9.5;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 ISHKDMQLGR 10
 11:1111
 DB 40 IEHKEIQGR 49

Search completed: February 27, 2002, 11:42:38
 Job time: 537 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: February 27, 2002, 11:33:41 ; Search time 281.76 seconds
(without alignments)
5.191 Million cell updates/sec

Title: US-09-446-109a-1
Perfect score: 52
Sequence: 1 ISHKMOLGR 10

Scoring table: GAPOP 10.0, Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

SPTREMBL_17:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	37	71.2	715	5	002330
2	35	67.3	178	2	09AKLO
3	35	67.3	178	2	09AKF6
4	35	67.3	319	2	09K8V8
5	35	67.3	512	10	09LUC8
6	35	67.3	833	10	09EF28
7	35	67.3	1489	2	09KE81
8	34	65.4	220	10	041367
9	34	65.4	363	1	09YD80
10	34	65.4	408	2	09K529
11	34	65.4	408	2	09AE72
12	34	65.4	413	11	09DA71
13	34	65.4	444	2	09ZFR3
14	34	65.4	591	3	09UT18
15	34	65.4	648	5	076763
16	34	65.4	649	5	061662
17	34	65.4	743	2	09KFS7
18	34	65.4	884	5	09VB71
19	34	65.4	896	12	090043

20	34	65.4	999	12	086519	086519 rice yellow
21	34	65.4	1217	12	09WS53	09WS53 simian t-ce
22	33	63.5	129	5	09XUL8	09XUL8 caenorhabdi
23	33	63.5	133	5	09VIC8	09VIC8 drosophila
24	33	63.5	136	5	09N2N8	09N2N8 styela plic
25	33	63.5	149	2	09L9U3	09L9U3 erwinia amy
26	33	63.5	291	2	P94824	P94824 helicobacte
27	33	63.5	291	2	025542	025542 helicobacte
28	33	63.5	291	2	050170	050170 helicobacte
29	33	63.5	292	2	09ZKX6	09ZKX6 helicobacte
30	33	63.5	293	2	09WZM1	09WZM1 thermotoga
31	33	63.5	312	2	09STO6	09STO6 yersinia ps
32	33	63.5	336	10	09ZPP5	09ZPP5 berberis st
33	33	63.5	1519	5	046015	046015 caenorhabdi
34	33	63.5	3187	11	063714	063714 ratius norv
35	32	61.5	92	11	099L12	099L12 mus musculu
36	32	61.5	126	6	09BE92	09BE92 macaca fasc
37	32	61.5	205	4	09BOK2	09BOK2 homo sapien
38	32	61.5	287	11	09D8T5	09D8T5 mus musculu
39	32	61.5	289	11	09CVH4	09CVH4 mus musculu
40	32	61.5	295	2	09K705	09K705 bacillus ha
41	32	61.5	337	2	09Z1I2	09Z1I2 streptococ
42	32	61.5	387	10	09L8B5	09L8B5 botrychium
43	32	61.5	395	2	09S513	09S513 loofah wic
44	32	61.5	512	10	09LUC9	09LUC9 arabidopsis
45	32	61.5	553	4	09UH74	09UH74 homo sapien

ALIGNMENTS

RESULT 1

ID 002330 PRELIMINARY: PRT: 715 AA.

AC 002330:

VT 01-JUL-1997 (TREMBLrel. 04, Created)

VT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)

VT 01-JAN-1999 (TREMBLrel. 09, Last annotation update)

DE T24D1.1 PROTEIN.

CN T24D1.1.

OS Caenorhabditis elegans.

UC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;

OC Rhabditidae; Pelodermidae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RA Cummings P.

RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Alnsough R., Anderson R., Baynes C., Berks M., Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A., Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L., Jones M., Kersey J., Kirsten J., Laister N., Latreille P., Lightning J., Lloyd C., McMurtry A., Mortimore B., O'Callaghan M., Parsons J., Percy C., Rinken L., Roopra A., Saunders D., Shownkeen R., Smalton N., Smith A., Sonhammer E., Staden R., Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R., Watson A., Weinstock L., Wilkinson-Spratt J., Woldman P., "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans.";

RT Nature 368:32-38(1994).

RL Nature 368:32-38(1994).

DR EMBL: Z81131; CAB03422.1;

SO SEQUENCE 715 AA; 81133 MW; 27C828513E5EE918 CRC64;

Query Match 71.2%; Score 37; DB 5; Length 715;
Best Local Similarity 66.7%; Pred. No. 19;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 226 SHEDELCR 234

RESULT 2

09AKL0 ID 09AKL0 PRELIMINARY: PRT: 178 AA.
 AC 09AKL0: 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE GUANOSINE-3,5-BIS(DIPHOSPHATE) 3-PYROPHOSPHOHYDROLASE.
 GN SPOTD.
 OS Rickettsia montana.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiidae; Rickettsia.
 RX NCBI_TaxID=33991;
 RN 111
 RP SEQUENCE FROM N.A.
 RA Andersson J.O., Andersson S.G.E.;
 RT "Pseudogenes, junk DNA and the Dynamics of Rickettsia genomes."
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ293330; CAC33644.1;
 KW Hydrolase.
 SQ SEQUENCE 178 AA: 21312 MW: 09433686BC10D16E CRC64;

Query Match 67.3%; Score 35; DB 2; Length 178;
 Best Local Similarity 75.0%; Pred. No. 12;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 HKDMOLCR 10
 111:111
 Db 95 HKDAELCR 102

RESULT 3

09AKF6 ID 09AKF6 PRELIMINARY: PRT: 178 AA.
 AC 09AKF6: 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE GUANOSINE-3,5-BIS(DIPHOSPHATE) 3-PYROPHOSPHOHYDROLASE.
 GN SPOTD.
 OS Rickettsia rickettsii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiidae; Rickettsia.
 RX NCBI_TaxID=783;
 RN 111
 RP SEQUENCE FROM N.A.
 RA STRAIN-84-21C;
 RA Andersson J.O., Andersson S.G.E.;
 RT "Pseudogenes, junk DNA and the dynamics of Rickettsia genomes."
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ293329; CAC33709.1;
 KW Hydrolase.
 SQ SEQUENCE 178 AA: 21359 MW: 1B9BF62B00039691 CRC64;

Query Match 67.3%; Score 35; DB 2; Length 178;
 Best Local Similarity 75.0%; Pred. No. 12;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 HKDMOLCR 10
 111:111
 Db 95 HKDAELCR 102

RESULT 4
 09K8V8 ID 09K8V8 PRELIMINARY: PRT: 319 AA.
 AC 09K8V8:

DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE ORNITHINE CARBAMOYLTRANSFERASE.
 GN ARGF OR BH2894.

OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 RX NCBI_TaxID=86665;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN-C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Masui N.,
 RA Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 RT halodurans and genomic sequence comparison with Bacillus subtilis."
 RL Nucleic Acids Res. 28:4317-4331(2000).
 DR EMBL: AP001517; BAB06613.1;
 DR InterPro: IPR002029; Carbmyltransf_asor.
 DR Pfam: PF00185; OTCace; 1.
 DR Pfam: PF02729; OTCace; N: 1.
 DR PRINTS: PR00100; AOTCACE.
 KW Transferase; Complete proteome.
 SQ SEQUENCE 319 AA: 34888 MW: D8EE2295DAED0A07 CRC64;

Query Match 67.3%; Score 35; DB 2; Length 319;
 Best Local Similarity 60.0%; Pred. No. 21;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 ISHKMOLCR 10
 111:111
 Db 84 LSPKDLQICR 93

RESULT 5

09LUC8 ID 09LUC8 PRELIMINARY: PRT: 512 AA.
 AC 09LUC8: 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE CYTOCHROME P450.
 DE CYTOCHROME P450.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 RX NCBI_TaxID=3702;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN 121
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RX MEDLINE=20277480; PubMed=10819329;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
 RT features of the regions of 4,504,864 bp covered by sixty pl and TAC
 RT clones."
 RL DNA Res. 7:131-135(2000).
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL: AB023038; BAB02398.1;
 DR InterPro: IPR001128; CYL_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 KW Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 512 AA: 58459 MW: D0AD22A9C907059D CRC64;

Query Match 67.3% Score 35; DB 10; Length 512;
 Best Local Similarity 85.7% Pred. No. 35;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 HKDMOLG 9
 111111
 DB 394 HKDMOLG 400

RESULT 6
 O9FF28 PRELIMINARY: PRT: 833 AA.

AC 09FF28; 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE DISEASE RESISTANCE PROTEIN-LIKE.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702.

RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
 RA Miyajima N., Tabata S.; PubMed-9330910;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
 RT features of the 1.6 Mb regions covered by twenty physically assigned
 RT pl clones."
 RT DNA Res. 4:215-230(1997).

DR EMBL: AB005248; BAB09346.1;
 DR InterPro: IPR001687; ATP_CTP_A.
 DR InterPro: IPR000767; Disease_Resist.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR002182; NB-ARC.
 DR Pfam: PF00560; LRR; 1.
 DR Pfam: PF00931; NB-ARC; 1.
 DR PRINTS: PR00364; DISEASEREST.
 DR SEQUENCE 833 AA; 95816 MW; FD7A8462803F5569 CRC64;

Query Match 67.3% Score 35; DB 10; Length 833;
 Best Local Similarity 75.0% Pred. No. 58;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 ISHKMOL 8
 11111111
 DB 112 INHKMOL 119

RESULT 7
 O9KE81 PRELIMINARY: PRT: 1489 AA.

AC 09KE81; 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE BH0975 PROTEIN.
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/staphylococcus group; Bacillus.
 OC NCBI_TaxID=86665;

RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
 RA Miyajima N., Tabata S.; PubMed-9330910;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
 RT features of the 1.6 Mb regions covered by twenty physically assigned
 RT pl clones."
 RT DNA Res. 4:215-230(1997).

RT halodurans and genomic sequence comparison with Bacillus subtilis."
 RT Nucleic Acids Res. 28:4317-4331(2000).
 DR EMBL: AP001510; BAB04694.1;
 DR InterPro: IPR002543; FtsK_SpoIIIE.
 DR Pfam: PF01580; FtsK_SpoIIIE; 2.
 KW Complete proteome.
 SQ SEQUENCE 1489 AA; 170265 MW; 48728BF8E5384625 CRC64;

Query Match 67.3% Score 35; DB 2; Length 1489;
 Best Local Similarity 75.0% Pred. No. 1;le+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 HKDMOLG 10
 11111111
 DB 1417 HKDVSIGR 1424

RESULT 8
 O41367 PRELIMINARY: PRT: 220 AA.

AC 041367; 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE 24 KDA RNA BINDING PROTEIN (FRAGMENT).
 OS Spinacia oleracea (Spinach).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.
 OC NCBI_TaxID=3562;

RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
 RA Miyajima N., Tabata S.; PubMed-9330910;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
 RT features of the 1.6 Mb regions covered by twenty physically assigned
 RT pl clones."
 RT DNA Res. 4:215-230(1997).

Query Match 65.4% Score 34; DB 10; Length 220;
 Best Local Similarity 77.8% Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 SHKMOLGR 10
 11111111
 DB 115 SHKMOLGR 123

RESULT 9
 O9YD80 PRELIMINARY: PRT: 363 AA.

AC 09YD80; 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
 DE HYPOTHETICAL 39.4 KDA PROTEIN APEI032.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Desulfurococcaceae;
 OC Aeropyrum.
 OC NCBI_TaxID=56636;

RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
 RA Miyajima N., Tabata S.; PubMed-9330910;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
 RT features of the 1.6 Mb regions covered by twenty physically assigned
 RT pl clones."
 RT DNA Res. 4:215-230(1997).

RP SEQUENCE FROM N.A.
 RC STRAIN-K1:
 RA MEDLINE=9310339; PubMed=10382966;
 RA Kawabayashi Y., Hino Y., Horikawa H., Yamazaki S., Halkawa Y.,
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankaï A., Kosugi H.,
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamaya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 RA Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.,
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 RT crenarchaeon, *Aeropyrum pernix* K1.";
 RL DNA Res. 6:83-101(1999).
 DR EMBL: AP000060; BAA80017.1; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 363 AA: 39437 MW: 67422FDD611901B7 CRC64;

Query Match 65.4%; Score 34; DB 1; Length 363;
 Best Local Similarity 60.0%; Pred. No.: 40;
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Caps 0;

OY 1 SHKMOQLGR 10
 : 1 1 1 1 1 1
 Db 286 VEHVDQLGR 295

RESULT 10
 O9K529 PRELIMINARY; PRT: 408 AA.
 AC O9K529:
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE FERREDOXIN REDUCTASE.
 GN BPAA.
 OS Pseudomonas sp. B4
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 NCBI_TaxID=59381;
 RN 11
 RC SEQUENCE FROM N.A.
 RP STRAIN-B4:
 RA Rodarte D., Willison J., Jounneau Y.,
 RT "Cloning, purification and molecular characterization of the biphenyl
 RT dioxygenase complex from *Pseudomonas* sp. strain B4.";
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases
 CC -1- COFACTOR: FAD (BY SIMILARITY).
 DR EMBL: AJ251217; CAB93969.1; -;
 DR InterPro: IPR000759; Adnrdx_reductase.
 DR InterPro: IPR001327; FAD_pyridox.
 DR InterPro: IPR000205; NAD_binding.
 DR InterPro: IPR001003; Pyridine_redox_2.
 DR InterPro: IPR001100; Pyridox_redox.
 DR InterPro: IPR003042; Rng_moxxygenase.
 DR Pfam: PF00070; Pyridox_redox_1.
 DR PRINTS: PR00419; ADXROTASE.
 DR PRINTS: PR00368; FADPNR.
 DR PRINTS: PR00411; PNDROTASE1.
 DR PRINTS: PR00469; PNDROTASE1.
 DR PRINTS: PR00420; RNCMNOXNAS.
 KW FAD; Flavoprotein; Oxidoreductase.
 SQ SEQUENCE 408 AA: 42911 MW: 81F3B10642D0688 CRC64;

Query Match 65.4%; Score 34; DB 2; Length 408;
 Best Local Similarity 66.7%; Pred. No.: 45;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Caps 0;

OY 2 SHKMOQLGR 10
 : 1 1 1 1 1 1
 Db 70 AHVDQLGR 78

RESULT 11
 O9AEY2 PRELIMINARY; PRT: 408 AA.
 AC O9AEY2:
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE FERREDOXIN REDUCTASE.
 GN BPBG.
 OS Pseudomonas sp. Cam-1
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 NCBI_TaxID=85672;
 RN 11
 RC SEQUENCE FROM N.A.
 RP STRAIN-CAM-1:
 RA Master E.R., Mohn W.W.,
 RT "Induction of bpha-encoding biphenyl dioxygenase in two
 RT Polychlorinated biphenyl-degrading bacteria, psychrotolerant species
 RT Cam-1 and mesophilic Burkholderia species LB400.";
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY027651; AKK14785.1; -;
 SQ SEQUENCE 408 AA: 42982 MW: 975E2009AA387979 CRC64;

Query Match 65.4%; Score 34; DB 2; Length 408;
 Best Local Similarity 66.7%; Pred. No.: 45;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Caps 0;

OY 2 SHKMOQLGR 10
 : 1 1 1 1 1 1
 Db 70 AHVDQLGR 78

RESULT 12
 O9DAC1 PRELIMINARY; PRT: 413 AA.
 AC O9DAC1:
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE 4933403M22RIK PROTEIN.
 GN 4933403M22RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN 11
 RC SEQUENCE FROM N.A.
 RP STRAIN-C57BL/6J; TISSUE=TESTIS;
 RA MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Atakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schirrali L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsch G.,
 RA Blake J., Boffelli D., Bojunga N., Carlini P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shilata Y., Storch K.-F.,
 RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whitlaker C., Wilming L.,
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL: AK016641; BAB30353.1; -;
 DR MGI: 1921321; 4933403M22RIK.

Search completed: February 27, 2002, 11:49:59
Job time: 978 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:33:41 : Search time 132.19 Seconds
(without alignments)
1.702 Million cell updates/sec

Title: US-09-446-109a-1
Perfect score: 52
Sequence: 1 ISHKDMQLGR 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*

- 1: /cgn2_6/ptodata/2/laa/5A.COMB.pep:*
- 2: /cgn2_6/ptodata/2/laa/5B.COMB.pep:*
- 3: /cgn2_6/ptodata/2/laa/6A.COMB.pep:*
- 4: /cgn2_6/ptodata/2/laa/6B.COMB.pep:*
- 5: /cgn2_6/ptodata/2/laa/PCPUS.COMB.pep:*
- 6: /cgn2_6/ptodata/2/laa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query Length	ID	Description
1	52	100.0	11	US-08-463-224-5	Sequence 5, Appl 1
2	52	100.0	11	US-08-463-377-5	Sequence 5, Appl 1
3	52	100.0	74	US-08-463-224-3	Sequence 3, Appl 1
4	52	100.0	74	US-08-463-377-3	Sequence 3, Appl 1
5	52	100.0	74	US-09-246-5008-10	Sequence 10, Appl 1
6	43	82.7	10	US-08-299-285-1	Sequence 1, Appl 1
7	43	82.7	10	US-08-463-224-52	Sequence 52, Appl 1
8	43	82.7	10	US-08-463-377-52	Sequence 52, Appl 1
9	43	82.7	10	US-08-985-126-1	Sequence 1, Appl 1
10	43	82.7	10	PCT-US95-11126-1	Sequence 1, Appl 1
11	39	75.0	10	US-08-299-285-2	Sequence 2, Appl 1
12	39	75.0	10	US-08-985-126-2	Sequence 2, Appl 1
13	39	75.0	10	PCT-US95-11126-2	Sequence 2, Appl 1
14	39	75.0	74	US-09-246-5008-11	Sequence 11, Appl 1
15	38	73.1	74	US-09-246-5008-12	Sequence 12, Appl 1
16	37	71.2	10	US-08-299-285-9	Sequence 9, Appl 1
17	37	71.2	10	US-08-985-126-9	Sequence 9, Appl 1
18	37	71.2	10	PCT-US95-11126-9	Sequence 9, Appl 1
19	34	65.4	77	US-09-246-5008-13	Sequence 13, Appl 1
20	33	63.5	10	US-08-299-285-5	Sequence 5, Appl 1
21	33	63.5	10	US-08-299-285-8	Sequence 8, Appl 1
22	33	63.5	10	US-08-299-285-26	Sequence 26, Appl 1
23	33	63.5	10	US-08-299-285-27	Sequence 27, Appl 1
24	33	63.5	10	US-08-299-285-28	Sequence 28, Appl 1
25	33	63.5	10	US-08-299-285-29	Sequence 29, Appl 1
26	33	63.5	10	US-08-299-285-30	Sequence 30, Appl 1
27	33	63.5	10	US-08-299-285-31	Sequence 31, Appl 1

28	33	63.5	10	US-08-299-285-32	Sequence 32, Appl 1
29	33	63.5	10	US-08-299-285-33	Sequence 33, Appl 1
30	33	63.5	10	US-08-299-285-34	Sequence 34, Appl 1
31	33	63.5	10	US-08-299-285-35	Sequence 35, Appl 1
32	33	63.5	10	US-08-299-285-36	Sequence 36, Appl 1
33	33	63.5	10	US-08-299-285-37	Sequence 37, Appl 1
34	33	63.5	10	US-08-985-126-5	Sequence 5, Appl 1
35	33	63.5	10	US-08-985-126-8	Sequence 8, Appl 1
36	33	63.5	10	US-08-985-126-26	Sequence 26, Appl 1
37	33	63.5	10	US-08-985-126-27	Sequence 27, Appl 1
38	33	63.5	10	US-08-985-126-28	Sequence 28, Appl 1
39	33	63.5	10	US-08-985-126-29	Sequence 29, Appl 1
40	33	63.5	10	US-08-985-126-30	Sequence 30, Appl 1
41	33	63.5	10	US-08-985-126-31	Sequence 31, Appl 1
42	33	63.5	10	US-08-985-126-32	Sequence 32, Appl 1
43	33	63.5	10	US-08-985-126-33	Sequence 33, Appl 1
44	33	63.5	10	US-08-985-126-34	Sequence 34, Appl 1
45	33	63.5	10	US-08-985-126-35	Sequence 35, Appl 1

ALIGNMENTS

RESULT 1
US-08-463-224-5
Sequence 5, Application US/08463224
Patent No. 5807824
GENERAL INFORMATION:
APPLICANT: Van Oostrum, Jan
APPLICANT: Boyar, William C.
APPLICANT: Galakatos, Nicholas G.
APPLICANT: Schmitz, Albert
APPLICANT: van Hecke, Gino
TITLE OF INVENTION: C5a Receptor Antagonists Having
NUMBER OF INVENTION: Substantially No. 5807824agonist Activity
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Lerner, David, Litzenberg, Krumholz & Mentlik
STREET: 600 South Avenue West
CITY: Westfield
STATE: NJ
COUNTRY: USA
ZIP: 07090
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,224
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Foley, Shawn P.
TELECOMMUNICATION INFORMATION:
TELEPHONE: 908-654-5000
TELEFAX: 908-654-7866
TELEX: 139-125
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 11 amino acids
TYPE: amino acid
TOPOLOGY: Linear
MOLECULE TYPE: peptide
US-08-463-224-5
Query Match 100.0%; Score 52; DB 1; Length 11;
Best Local Similarity 100.0%; Pred. No. 5; Se-05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Caps 0;
|||||||||
QY 1 ISHKDMQLGR 10

Db 2 ISHKMOLGR 11

RESULT 2

US-08-463-377-5
Sequence 5, Application US/08463377
Patent No. 5837499

GENERAL INFORMATION:

APPLICANT: van Oostrum, Jan
APPLICANT: Boyar, William C.
APPLICANT: Galakatos, Nicholas G.
APPLICANT: Schmitz, Albert

APPLICANT: van Heeke, Gino

TITLE OF INVENTION: C5a Receptor Antagonists Having

NUMBER OF SEQUENCES: 67 Substantially No. 5837499agonist Activity

CORRESPONDENCE ADDRESS:

ADDRESSEE: Lerner, David, Littenberg, Krumholz & Mentlik
STREET: 600 South Avenue West
CITY: Westfield

STATE: NJ

COUNTRY: USA

ZIP: 07090

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/463,377

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Foley, Shawn P.

TELECOMMUNICATION INFORMATION:

TELEPHONE: 908-654-5000

TELEFAX: 908-654-7866

TELEX: 139-125

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 11 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-463-377-5

Query Match 100.0%; Score 52; DB 2; Length 11;

Best Local Similarity 100.0%; Pred. No. 5.5e-05;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISHKMOLGR 10

Db 2 ISHKMOLGR 11

RESULT 3

US-08-463-224-3

Sequence 3, Application US/08463224

Patent No. 5807824

GENERAL INFORMATION:

APPLICANT: van Oostrum, Jan

APPLICANT: Boyar, William C.

APPLICANT: Galakatos, Nicholas G.

APPLICANT: Schmitz, Albert

APPLICANT: van Heeke, Gino

TITLE OF INVENTION: C5a Receptor Antagonists Having

NUMBER OF SEQUENCES: 67 Substantially No. 5807824agonist Activity

CORRESPONDENCE ADDRESS:

ADDRESSEE: Lerner, David, Littenberg, Krumholz & Mentlik

STREET: 600 South Avenue West

CITY: Westfield

STATE: NJ

COUNTRY: USA

ZIP: 07090

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/463,224

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Foley, Shawn P.

TELECOMMUNICATION INFORMATION:

TELEPHONE: 908-654-5000

TELEFAX: 908-654-7866

TELEX: 139-125

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 74 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-463-224-3

Query Match 100.0%; Score 52; DB 1; Length 74;

Best Local Similarity 100.0%; Pred. No. 0.00045;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ISHKMOLGR 10

Db 65 ISHKMOLGR 74

RESULT 4

US-08-463-377-3

Sequence 3, Application US/08463377

Patent No. 5837499

GENERAL INFORMATION:

APPLICANT: van Oostrum, Jan

APPLICANT: Boyar, William C.

APPLICANT: Galakatos, Nicholas G.

APPLICANT: Schmitz, Albert

APPLICANT: van Heeke, Gino

TITLE OF INVENTION: C5a Receptor Antagonists Having

NUMBER OF SEQUENCES: 67 Substantially No. 5837499agonist Activity

CORRESPONDENCE ADDRESS:

ADDRESSEE: Lerner, David, Littenberg, Krumholz & Mentlik

STREET: 600 South Avenue West

CITY: Westfield

STATE: NJ

COUNTRY: USA

ZIP: 07090

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/463,377

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Foley, Shawn P.

TELECOMMUNICATION INFORMATION:

TELEPHONE: 908-654-5000

TELEFAX: 908-654-7866

TELEX: 139-125

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 74 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-463-377-3

Query Match 100.0%; Score 52; DB 2; Length 74;
Best Local Similarity 100.0%; Pred. No. 0.00045;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ISHKMOLGR 10
|||||
Db 65 ISHKMOLGR 74

RESULT 5
US-09-246-500B-10
Sequence 10, Application US/09246500B
Patent No. 6235494
GENERAL INFORMATION:
APPLICANT: Huggli, Tony E.
TITLE OF INVENTION: Substrates for Assessing Mannan-Binding
TITLE OF INVENTION: Protein-Associated Serine Protease Activity and Methods
FILE REFERENCE: 24730-2204
CURRENT APPLICATION NUMBER: US/09/246,500B
CURRENT FILING DATE: 1999-02-08
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10
LENGTH: 74
TYPE: PRP
ORGANISM: Human C5a Anaphylatoxin
FEATURE:
NAME/KEY: CARBOHYD
LOCATION: (64)...(66)
US-09-246-500B-10

Query Match 100.0%; Score 52; DB 4; Length 74;
Best Local Similarity 100.0%; Pred. No. 0.00045;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ISHKMOLGR 10
|||||
Db 65 ISHKMOLGR 74

RESULT 6
US-08-299-285-1
Sequence 1, Application US/08299285
Patent No. 5696230
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
US-08-299-285-1

Query Match 82.7%; Score 43; DB 1; Length 10;
Best Local Similarity 90.0%; Pred. No. 0.0039;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ISHKMOLGR 10
|||||
Db 1 ISFKDMOLGR 10

RESULT 7
US-08-463-224-52
Sequence 52, Application US/08463224
Patent No. 5807824
GENERAL INFORMATION:
APPLICANT: van Oostrum, Jan
APPLICANT: Boyar, William C.
APPLICANT: Galakatos, Nicholas G.
APPLICANT: Schmitz, Albert
APPLICANT: van Heeke, Gino
TITLE OF INVENTION: C5a Receptor Antagonists Having
TITLE OF INVENTION: Substantially No. 5807824Antagonist Activity
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Lerner, David, Litlenberg, Krumholz & Mentlik
STREET: 600 South Avenue West
CITY: Westfield
STATE: NJ
COUNTRY: USA
ZIP: 07090
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,224
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Foley, Shawn P.
TELECOMMUNICATION INFORMATION:
TELEPHONE: 908-654-5000
TELEFAX: 908-654-7866
TELEX: 139-125
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: peptide
US-08-463-224-52

Query Match 82.7%: Score 43; DB 1; Length 10;
Best Local Similarity 90.0%: Pred. No. 0.0039;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ISHKDQQLGR 10
1111111111
Db 1 ISFKDQQLGR -10

RESULT 8
US-08-463-377-52
Sequence 52, Application US/08463377
Patent No. 5837499
GENERAL INFORMATION:
APPLICANT: van Oostrum, Jan
APPLICANT: Boyar, William C.
APPLICANT: Galakatos, Nicholas G.
APPLICANT: Schmitz, Albert
APPLICANT: van Heeke, Gino
TITLE OF INVENTION: Csa Receptor Antagonists Having
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lerner, David, Littenberg, Krumholz & Menclik
STREET: 600 South Avenue West
CITY: Westfield
STATE: NJ
COUNTRY: USA
ZIP: 07090
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463.377
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Foley, Shawn P.
TELECOMMUNICATION INFORMATION:
TELEPHONE: 908-654-5000
TELEFAX: 908-654-7866
TELEX: 139-125
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-463-377-52

Query Match 82.7%: Score 43; DB 2; Length 10;
Best Local Similarity 90.0%: Pred. No. 0.0039;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ISHKDQQLGR 10
1111111111
Db 1 ISFKDQQLGR 10

RESULT 9
US-08-965-126-1
Sequence 1, Application US/08985126
Patent No. 5942599
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.

APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
TITLE OF INVENTION: C-Terminal Analogs of Csa Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/985.126
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/299.285
FILING DATE: 31-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: NO, 5942599 Relevantant
TOPOLOGY: NO, 5942599 Relevantant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
US-08-985-126-1

Query Match 82.7%: Score 43; DB 2; Length 10;
Best Local Similarity 90.0%: Pred. No. 0.0039;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ISHKDQQLGR 10
1111111111
Db 1 ISFKDQQLGR 10

RESULT 10
PCT-US95-11126-1
Sequence 1, Application PC/TUS9511126
GENERAL INFORMATION:

APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-selective
TITLE OF INVENTION: C-Terminal Analogs of Csa Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman, P.C.
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/11126
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/299,285
FILING DATE: 31-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
PCT-US95-11126-1

Query Match 82.7%; Score 43; DB 5; Length 10;
Best Local Similarity 90.0%; Pred. No. 0.0039;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ISHKDQLGR 10
1111111111
Db 1 ISFKDQLGR 10

RESULT 11
US-08-299-285-2
Sequence 2, Application US/08299285
Patent No. 5696230
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100

TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
US-08-299-285-2

Query Match 75.0%; Score 39; DB 1; Length 10;
Best Local Similarity 88.9%; Pred. No. 0.027;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 SHKDMQLGR 10
1111111111
Db 2 SFKDMQLGR 10

RESULT 12
US-08-985-126-2
Sequence 2, Application US/08985126
Patent No. 5942599
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/985,126
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: No. 5942599 Relevant
TOPOLOGY: No. 5942599 Relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
US-08-985-126-2

Query Match 75.0%; Score 39; DB 2; Length 10;
Best Local Similarity 88.9%; Pred. No. 0.027;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 SHKDMOLGR 10
1 11111111
DB 2 SFKDMOLGR 10

RESULT 13

PCT-US95-11126-2
Sequence 2; Application PCT/US9511126
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman, P.C.
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/11126
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/299,285
FILING DATE: 31-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO. 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:
NAME/KEY: Modified-site
LOCATION: 9..10
OTHER INFORMATION: /product="D-Alanine; N-methylated
PCT-US95-11126-2

Query Match 75.0%; Score 39; DB 5; Length 10;
Best Local Similarity 88.9%; Pred. No. 0.027;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 SHKDMOLGR 10
1 11111111
DB 2 SFKDMOLGR 10

RESULT 14
US-09-246-500B-11
Sequence 11; Application US/09246500B
Patent No. 6235494
GENERAL INFORMATION:
APPLICANT: Hugli, Tony E.
TITLE OF INVENTION: Substrates for Assessing Mannan-Binding
TITLE OF INVENTION: Protein-Associated Serine Protease Activity and Methods
FILE REFERENCE: 24730-2204
CURRENT APPLICATION NUMBER: US/09/246,500B
CURRENT FILING DATE: 1999-02-08
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11
LENGTH: 74
TYPE: PRT
ORGANISM: Pig C5a Anaphylatoxin
FEATURE:
NAME/KEY: CARBOHYD
LOCATION: (64)...(66)
US-09-246-500B-11

Query Match 75.0%; Score 39; DB 4; Length 74;
Best Local Similarity 77.8%; Pred. No. 0.25;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 SHKDMOLGR 10
11111111
DB 66 SHKNIOLGR 74

RESULT 15
US-09-246-500B-12
Sequence 12; Application US/09246500B
Patent No. 6235494
GENERAL INFORMATION:
APPLICANT: Hugli, Tony E.
TITLE OF INVENTION: Substrates for Assessing Mannan-Binding
TITLE OF INVENTION: Protein-Associated Serine Protease Activity and Methods
FILE REFERENCE: 24730-2204
CURRENT APPLICATION NUMBER: US/09/246,500B
CURRENT FILING DATE: 1999-02-08
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12
LENGTH: 74
TYPE: PRT
ORGANISM: Cow C5a Anaphylatoxin
US-09-246-500B-12

Query Match 73.1%; Score 38; DB 4; Length 74;
Best Local Similarity 70.0%; Pred. No. 0.4;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 ISHKDMOLGR 10
1 11111111
DB 65 IHKNIOLGR 74

Search completed: February 27, 2002, 11:36:02
Job time: 141 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: February 27, 2002, 11:41:12 ; Search time 303.5 Seconds
(without alignments)
2.441 Million cell updates/sec

Title: US-09-446-109a-2
Perfect score: 53
Sequence: 1 YSEKDWQLGR 10

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_1101.*
1: /SIDSB/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSB/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSB/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSB/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSB/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSB/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSB/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSB/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSB/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSB/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSB/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSB/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSB/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSB/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSB/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSB/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSB/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSB/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSB/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSB/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSB/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDSB/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
1	53	100.0	10	17	AA894465 C5a anaphylatoxin
2	51	96.2	10	17	AA894472 C5a anaphylatoxin
3	47	88.7	10	17	AA894489 C5a anaphylatoxin
4	47	88.7	10	17	AA894490 C5a anaphylatoxin
5	47	88.7	10	17	AA894492 C5a anaphylatoxin
6	47	88.7	10	17	AA894493 C5a anaphylatoxin
7	47	88.7	10	17	AA894494 C5a anaphylatoxin
8	47	88.7	10	17	AA894495 C5a anaphylatoxin
9	47	88.7	10	17	AA894496 C5a analogue C5a(6
10	47	88.7	10	17	AA894497 C5a analogue C5a(6
11	47	88.7	10	17	AA894498 C5a analogue C5a(6

12	47	88.7	10	17	AA894499 C5a analogue C5a(6
13	47	88.7	10	17	AA894500 C5a analogue C5a(6
14	47	88.7	10	17	AA894471 C5a anaphylatoxin
15	47	88.7	10	17	AA894468 C5a anaphylatoxin
16	46	86.8	10	16	AA875501 Control decapetide
17	46	86.8	10	17	AA894488 C5a anaphylatoxin
18	46	86.8	10	17	AA894464 Human C5a anaphyla
19	46	86.8	10	17	AA894467 C5a anaphylatoxin
20	45	84.9	10	17	AA894466 C5a anaphylatoxin
21	42	79.2	10	17	AA894474 C5a anaphylatoxin
22	41	77.4	10	17	AA894470 C5a anaphylatoxin
23	41	77.4	10	17	AA894473 C5a anaphylatoxin
24	41	77.4	10	17	AA894478 C5a anaphylatoxin
25	41	77.4	10	17	AA894479 C5a anaphylatoxin
26	40	75.5	10	17	AA894486 C5a anaphylatoxin
27	40	75.5	10	17	AA894487 C5a anaphylatoxin
28	39	73.6	10	17	AA894485 C5a anaphylatoxin
29	39	73.6	10	17	AA894477 C5a anaphylatoxin
30	39	73.6	10	17	AA894483 C5a anaphylatoxin
31	39	73.6	11	16	AA875500 C5a(64-74) C-termi
32	39	73.6	20	22	AA874056 Human C5a peptide
33	39	73.6	74	8	AA875166 Human anaphylatoxi
34	39	73.6	74	16	AA875497 Human C5a protein.
35	39	73.6	74	20	AA89580 Solid phase sequen
36	39	73.6	74	22	AA874053 Human C5a anaphyla
37	39	73.6	74	22	AA874119 Human C5a. Homo s
38	39	73.6	74	22	AA874119 Variant human C5a.
39	39	73.6	1676	16	AA877604 Pro-C5 polypeptide
40	38	71.7	10	17	AA894476 C5a anaphylatoxin
41	38	71.7	384	21	AA14386 Zee mays RecA-like
42	36	67.9	10	17	AA894475 C5a anaphylatoxin
43	36	67.9	10	17	AA894484 C5a anaphylatoxin
44	36	67.9	441	22	AAU12284 Human PKO593 poly
45	36	67.9	441	22	AA868888 Human RBCAP polype

ALIGNMENTS

RESULT 1

ID AA894465 standard: peptide: 10 AA.

AC AA894465;

DT 17-OCT-1996 (first entry)

C5a anaphylatoxin C-terminal analogue C5a(65-74) Tyr65.

C-terminal: human: C5a anaphylatoxin; analogue: beta-turn;
constrained backbone conformation: proinflammatory response;
spasmogenesis: platelet aggregation: neutrophil: non-mediated;
mediated: increase: cell membrane: vascular: permeability;
neutrophil polarisation: neutrophil enzyme release: treatment;
immune adjuvant: immunodeficiency: augmentation: immune therapy;
cancer: drug delivery: CNS disorder: central nervous system;
Alzheimer's disease: tumour: high affinity: receptor antagonist;
development: non-steroidal: anti-inflammatory.

Homio sapiens.

Key Location/Qualifiers

Region 4..7 /note="beta-turn"

W09606629-A1.

07-MAR-1996.

31-AUG-1995; 95WO-US11126.

31-AUG-1994; 94US-0299285.

PA (UYNE-) UNIV NEBRASKA.
 PA (UYOU) UNIV QUEENSLAND.
 PI Klrnarsky L, Sanderson DG, Sherman SA, Taylor SM;
 DR WPI: 1996-160140/16.
 XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
 PT spasmogenesis, platelet aggregation and increases in cell membrane
 PT permeability
 XX
 PS Claim 8: Page 77: 116pp: English.
 CC The present peptide is an analogue of a C-terminal decapeptide
 CC region from human C5a anaphylatoxin, which has a constrained
 CC backbone conformation (compared to the natural peptide's flexible
 CC C-terminal region) comprising a beta-turn. Analogues with type
 CC (II or V) and type III beta-turns elicit proinflammatory responses
 CC characterised by spasmogenesis, platelet aggregation and neutrophil
 CC non-mediated increases of cell membrane (specifically vascular)
 CC permeability, or neutrophil polarisation, neutrophil enzyme
 CC release and neutrophil mediated increases in cell membrane (esp.
 CC vascular) permeability, respectively.
 CC The analogues can be used as immune adjuvants for the treatment of
 CC immune deficiency disorders, and for augmenting standard immune
 CC therapy for the treatment of cancer, without inflammatory side
 CC effects. They can also be used to facilitate drug delivery by
 CC increasing vascular permeability, e.g. in the treatment of CNS
 CC disorders such as Alzheimer's disease or tumours, and to develop
 CC high affinity C5a receptor antagonists useful as non-steroidal
 CC anti-inflammatory agents.
 CC In a type (II or V) beta-turn specific human foetal artery smooth
 CC muscle contraction assay, and a type III beta-turn specific
 CC neutrophil polarisation assay the peptide had respective EC50
 CC (microm) activities of 11.2 and 4.0, compared to 0.018 and 0.0013
 CC for C5a.
 XX
 XX Sequence 10 AA:
 SO
 Query Match 100.0%; Score 53; DB 17; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.00014;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 YSFKDMLGR 10
 Db 1 ysfkdmqlgr 10
 RESULT 2
 AAR94472
 ID AAR94472 standard; peptide: 10 AA.
 AC AAR94472:
 XX 17-OCT-1996 (first entry)
 DE C5a anaphylatoxin C-terminal analogue C5a(65-74) Tyr65 Ile72.
 XX
 KW C-terminal: human; C5a anaphylatoxin; analogue; beta-turn;
 KW constrained backbone conformation; proinflammatory response;
 KW spasmogenesis; platelet aggregation; neutrophil; non-mediated;
 KW mediated; increase; cell membrane; vascular; permeability;
 KW neutrophil polarisation; neutrophil enzyme release; treatment;
 KW immune adjuvant; immunodeficiency; augmentation; immune therapy;
 KW cancer; drug delivery; CNS disorder; central nervous system;
 KW Alzheimer's disease; tumour; high affinity; receptor antagonist;
 KW development; non-steroidal; anti-inflammatory.
 KW
 XX Homo sapiens.
 OS
 XX Key Location/Qualifiers
 FII Region 4..7
 FT

FT /note= "beta-turn"
 XX
 XX MO9606629-A1.
 XX
 PD 07-MAR-1996.
 XX
 XX 31-AUG-1995; 95WO-US11126.
 XX
 XX 31-AUG-1994; 94US-0299285.
 XX
 XX (UYNE-) UNIV NEBRASKA.
 PA (UYOU) UNIV QUEENSLAND.
 PI Klrnarsky L, Sanderson DG, Sherman SA, Taylor SM;
 DR WPI: 1996-160140/16.
 XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
 PT spasmogenesis, platelet aggregation and increases in cell membrane
 PT permeability
 XX
 PS Example 2: Page 80: 116pp: English.
 CC The present peptide is an analogue of a C-terminal decapeptide
 CC region from human C5a anaphylatoxin, which has a constrained
 CC backbone conformation (compared to the natural peptide's flexible
 CC C-terminal region) comprising a beta-turn. Analogues with type
 CC (II or V) and type III beta-turns elicit proinflammatory responses
 CC characterised by spasmogenesis, platelet aggregation and neutrophil
 CC non-mediated increases of cell membrane (specifically vascular)
 CC permeability, or neutrophil polarisation, neutrophil enzyme
 CC release and neutrophil mediated increases in cell membrane (esp.
 CC vascular) permeability, respectively.
 CC The analogues can be used as immune adjuvants for the treatment of
 CC immune deficiency disorders, and for augmenting standard immune
 CC therapy for the treatment of cancer, without inflammatory side
 CC effects. They can also be used to facilitate drug delivery by
 CC increasing vascular permeability, e.g. in the treatment of CNS
 CC disorders such as Alzheimer's disease or tumours, and to develop
 CC high affinity C5a receptor antagonists useful as non-steroidal
 CC anti-inflammatory agents.
 CC In a type (II or V) beta-turn specific human foetal artery smooth
 CC muscle contraction assay, and a type III beta-turn specific
 CC neutrophil polarisation assay the peptide had respective EC50
 CC (microm) activities of 286 and 3.7, compared to 0.018 and 0.0013
 CC for C5a.
 XX
 XX Sequence 10 AA:
 SO
 Query Match 96.2%; Score 51; DB 17; Length 10;
 Best Local Similarity 90.0%; Pred. No. 0.00036;
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 YSFKDMLGR 10
 Db 1 ysfkdmqlgr 10
 RESULT 3
 AAR94489
 ID AAR94489 standard; peptide: 10 AA.
 AC AAR94489:
 XX 17-OCT-1996 (first entry)
 DE C5a anaphylatoxin analogue C5a(65-74) Tyr65 D-Ala73 NMe-Arg10.
 XX
 KW C-terminal: human; C5a anaphylatoxin; analogue; beta-turn;
 KW constrained backbone conformation; proinflammatory response;
 KW spasmogenesis; platelet aggregation; neutrophil; non-mediated;
 KW mediated; increase; cell membrane; vascular; permeability;
 KW

XX	neutrophil polarisation: neutrophil enzyme release; treatment;
KM	immune adjuvant: immunodeficiency; augmentation; immune therapy;
KM	cancer; drug delivery; CNS disorder; central nervous system;
KM	Alzheimer's disease; tumour; high affinity; receptor antagonist;
KM	development; non-steroidal; anti-inflammatory.
XX	
OS	Homo sapiens.
XX	
FT	Key
FT	Region
FT	4..7
FT	/note= "beta-turn"
FT	Misc-difference
FT	9
FT	/note= "D-form residue"
FT	Modified-site
FT	10
FT	/note= "N-methylated arganine"
XX	
PN	W09606629-A1.
XX	
PD	07-MAR-1996.
XX	
PE	31-AUG-1995:
XX	95WO-US11126.
XX	
IR	31-AUG-1994:
XX	94US-0299285.
XX	
PA	(UYNE-) UNIV NEBRASKA.
PA	(UYUN-) UNIV QUEBEC/STAN.
PI	
PI	Kliratsky L, Sanderson DG, Sherman SA, Taylor SM;
DR	WPI: 1996-160140/16.
XX	
PT	C-terminal analogues of C5a anaphylatoxin - induce e.g.
PT	spasmogenesis, platelet aggregation and increases in cell membrane
PT	permeability
XX	
PS	Example 5: Page 89; 116pp: English.
XX	
CC	The present peptide is an analogue of a C-terminal decapeptide
CC	region from human C5a anaphylatoxin, which has a constrained
CC	backbone conformation (compared to the natural peptide's flexible-
CC	C-terminal region) comprising a beta-turn. Analogues with type
CC	(II or V) and type III beta-turns elicit proinflammatory responses
CC	characterised by spasmogenesis, platelet aggregation and neutrophil
CC	non-mediated increases of cell membrane (specifically vascular)
CC	permeability, or neutrophil polarisation, neutrophil enzyme
CC	release and neutrophil mediated increases in cell membrane (esp.
CC	vascular) permeability, respectively.
CC	The analogues can be used as immune adjuvants for the treatment of
CC	immune deficiency disorders, and for augmenting standard immune
CC	therapy for the treatment of cancer, without inflammatory side
CC	effects. They can also be used to facilitate drug delivery by
CC	increasing vascular permeability, e.g. in the treatment of CNS
CC	disorders such as Alzheimer's disease or tumours, and to develop
CC	high affinity C5a receptor antagonists useful as non-steroidal
CC	anti-inflammatory agents.
XX	
Sequence	10 AA:
10	

```

Query Match      88.7%: Score 47: DB 17: Length 10:
Best Local Similarity 90.0%: Pred. No. 0.0024:
Matches 9: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

QY      1 YSPKDMQLGR 10
          |||||
          1
Db       1 YSFKDMQLGR 10

RESULT 4
AAK94490
ID AAK94490 standard: peptide: 10 AA.
XX
XX AAK94490:

```

XX	17-OCT-1996	(first entry)
DT	C5a anaphylatoxin analogue C5a(65-74) Tyr65 alphaMe-Leu72 D-Ala73.	
XX		
XZ	C-terminal; human; C5a anaphylatoxin; analogue; beta-turn;	
KW	constrained backbone conformation; proinflammatory response;	
KW	spasmogenesis; platelet aggregation; neutrophil; non-mediated;	
KW	mediated; increase; cell membrane; vascular; permeability;	
KW	neutrophil polarisation; neutrophil enzyme release; treatment;	
KW	immune adjuvant; immunodeficiency; augmentation; immune therapy;	
KW	cancer; drug delivery; CNS disorder; central nervous system;	
KW	Alzheimer's disease; tumour; high affinity; receptor antagonist;	
KW	development; non-steroidal; anti-inflammatory.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	
FT	Region	
FT	/note= "beta-turn"	
FT	Misc-difference	
FT	/note= "D-form residue"	
FT	Modified-site	
FT	/note= "alpha-methylated leucine"	
XX		
IN	WO9606629-A).	
XX		
PD	07-MAR-1996.	
XX		
PF	31-AUG-1995; 95WO-US11126.	
XX		
PR	31-AUG-1994; 94US-0299285.	
XX		
PA	(UYNE-) UNIV NEBRASKA.	
PA	(UYNU) UNIV QUEBENSLAND.	
XX		
PI	Klirnarsky L, Sanderson DG, Sherman SA, Taylor SK:	
DR	WPI: 1996-160140/16.	
XX		
PT	C-terminal analogues of C5a anaphylatoxin - induce e.g.	
PT	spasmogenesis, platelet aggregation and increases in cell membrane	
PT	permeability	
XX		
PS	Example 5; Page 89; 116pp; English.	
XX		
CC	The present peptide is an analogue of a C-terminal decapeptide	
CC	region from human C5a anaphylatoxin, which has a constrained	
CC	backbone conformation (compared to the natural peptide's flexible	
CC	C-terminal region) comprising a beta-turn. Analogues with type	
CC	(II or V) and type III beta-turns elicit proinflammatory responses	
CC	characterised by spasmogenesis, platelet aggregation and neutrophil	
CC	non-mediated increases of cell membrane (specifically vascular)	
CC	permeability, or neutrophil polarisation, neutrophil enzyme	
CC	release and neutrophil mediated increases in cell membrane (esp.	
CC	vascular) permeability, respectively	
CC	The analogues can be used as immune adjuvants for the treatment of	
CC	immune deficiency disorders, and for augmenting standard immune	
CC	therapy for the treatment of cancer, without inflammatory side	
CC	effects. They can also be used to facilitate drug delivery by	
CC	increasing vascular permeability, e.g. in the treatment of CNS	
CC	disorders such as Alzheimer's disease or tumours, and to develop	
CC	high affinity C5a receptor antagonists useful as non-steroidal	
CC	anti-inflammatory agents.	
XX		
SQ	Sequence 10 AA:	
XX		

Query Match	88.7%	Score 47	DB 17	length 10
Best Local Similarity	90.0%	Pred No. 0.0024		
Matches 9	Conservative	0	Mismatches 1	Indels 0
				Gaps 0
QY	1	YSFKDMOLGR	10	

Db 1 ysfkdmqlar 10

RESULT 5
AAR94492
ID AAR94492 standard: peptide: 10 AA.
AC AAR94492;
DT 17-OCT-1996 (first entry)
XX C5a anaphylatoxin analogue C5a(65-74) Tyr65 NMe-D-Ala73.
DE
XX
XX C-terminal: human: C5a anaphylatoxin; analogue: beta-turn;
KW constrained backbone conformation; proinflammatory response;
KW spasmogenesis; platelet aggregation; neutrophil: non-mediated;
KW mediated: increase; cell membrane: vascular: permeability;
KW neutrophil polarisation; neutrophil enzyme release; treatment;
KW immune adjuvant; immunodeficiency; augmentation; immune therapy;
KW cancer; drug delivery; CNS disorder; central nervous system;
KW Alzheimer's disease; tumour; high affinity; receptor antagonist;
KW development; non-steroidal; anti-inflammatory.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH Region 4..7
FT /note= "beta-turn"
FT Modified-site 9
FT /note= "N-methylated D-alanine"
XX
XX W09606629-A1.
XX 07-MAR-1996.
XX
XX 31-AUG-1995: 95WO-US11126.
XX
XX 31-AUG-1994: 94US-0299285.
XX
XX (UYNE-) UNIV NEBRASKA.
PA (UYOU) UNIV QUEENSLAND.
XX
XX Kirmarsky L, Sanderson DG, Sherman SA, Taylor SM;
PI
XX
XX WPI: 1996-160140/16.
XX
XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
PT spasmogenesis, platelet aggregation and increases in cell membrane
PT permeability.
XX
XX Example 5: Page 90: 116pp: English.
XX
XX The present peptide is an analogue of a C-terminal decapeptide
CC region from human C5a anaphylatoxin, which has a constrained
CC backbone conformation (compared to the natural peptide's flexible
CC C-terminal region) comprising a beta-turn. Analogues with type
CC (II or V) and type I/II beta-turns elicit proinflammatory responses
CC characterised by spasmogenesis, platelet aggregation and neutrophil
CC non-mediated increases of cell membrane (specifically vascular)
CC permeability, or neutrophil polarisation, neutrophil enzyme
CC release and neutrophil mediated increases in cell membrane (esp.
CC vascular) permeability, respectively.
CC The analogues can be used as immune adjuvants for the treatment of
CC immune deficiency disorders, and for augmenting standard immune
CC therapy for the treatment of cancer, without inflammatory side
CC effects. They can also be used to facilitate drug delivery by
CC increasing vascular permeability, e.g. in the treatment of CNS
CC disorders such as Alzheimer's disease or tumours, and to develop
CC high affinity C5a receptor antagonists useful as non-steroidal
CC anti-inflammatory agents.
XX
XX Sequence 10 AA:
SO

Query Match 88.7%; Score 47; DB 17; Length 10;
Best Local Similarity 90.0%; Pred. No. 0.0024;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Gy 1 ysfkdmqlar 10
Db 1 ysfkdmqlar 10

RESULT 6
AAR94493
ID AAR94493 standard: peptide: 10 AA.
AC AAR94493;
DT 17-OCT-1996 (first entry)
XX C5a anaphylatoxin analogue C5a(65-74) Tyr65 NMe-Leu72 D-Ala73.
DE
XX
XX C-terminal: human: C5a anaphylatoxin; analogue: beta-turn;
KW constrained backbone conformation; proinflammatory response;
KW spasmogenesis; platelet aggregation; neutrophil: non-mediated;
KW mediated: increase; cell membrane: vascular: permeability;
KW neutrophil polarisation; neutrophil enzyme release; treatment;
KW immune adjuvant; immunodeficiency; augmentation; immune therapy;
KW cancer; drug delivery; CNS disorder; central nervous system;
KW Alzheimer's disease; tumour; high affinity; receptor antagonist;
KW development; non-steroidal; anti-inflammatory.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH Region 4..7
FT /note= "beta-turn"
FT Misc-difference 9
FT /note= "D-form residue"
FT Modified-site 8
FT /note= "N-methylated leucine"
XX
XX W09606629-A1.
XX 07-MAR-1996.
XX
XX 31-AUG-1995: 95WO-US11126.
XX
XX 31-AUG-1994: 94US-0299285.
XX
XX (UYNE-) UNIV NEBRASKA.
PA (UYOU) UNIV QUEENSLAND.
XX
XX Kirmarsky L, Sanderson DG, Sherman SA, Taylor SM;
PI
XX
XX WPI: 1996-160140/16.
XX
XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
PT spasmogenesis, platelet aggregation and increases in cell membrane
PT permeability.
XX
XX Example 5: Page 91: 116pp: English.
XX
XX The present peptide is an analogue of a C-terminal decapeptide
CC region from human C5a anaphylatoxin, which has a constrained
CC backbone conformation (compared to the natural peptide's flexible
CC C-terminal region) comprising a beta-turn. Analogues with type
CC (II or V) and type I/II beta-turns elicit proinflammatory responses
CC characterised by spasmogenesis, platelet aggregation and neutrophil
CC non-mediated increases of cell membrane (specifically vascular)
CC permeability, or neutrophil polarisation, neutrophil enzyme
CC release and neutrophil mediated increases in cell membrane (esp.
CC vascular) permeability, respectively.
CC The analogues can be used as immune adjuvants for the treatment of

CC immune deficiency disorders, and for augmenting standard immune
 CC therapy for the treatment of cancer, without inflammatory side
 CC effects. They can also be used to facilitate drug delivery by
 CC increasing vascular permeability, e.g. in the treatment of CNS
 CC disorders such as Alzheimer's disease or tumours, and to develop
 CC high affinity C5a receptor antagonists useful as non-steroidal
 CC anti-inflammatory agents.

XX
 CC Sequence 10 AA:

SO

Query Match 88.7%: Score 47; DB 17; Length 10;
 Best Local Similarity 90.0%; Pred. No. 0.0024;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSFKDMQLGR 10
 |||||
 Db 1 ysfkdmqlar 10

RESULT 7
 AAR94494
 ID AAR94494 standard; peptide: 10 AA.
 AC AAR94494;
 XX
 DT 17-OCT-1996 (first entry)

C5a anaphylatoxin analogue C5a(65-74) Tyr65 NME-Gln71 D-Ala73.

XX
 KW C-terminal: human; C5a anaphylatoxin; analogue: beta-turn;
 KW constrained backbone conformation: proinflammatory response;
 KW spasmogenesis: platelet aggregation; neutrophil: non-mediated;
 KW mediated; increase; cell membrane; vascular; permeability;
 KW neutrophil polarisation; neutrophil enzyme release; treatment;
 KW immune adjuvant; immunodeficiency; augmentation; immune therapy;
 KW cancer; drug delivery; CNS disorder; central nervous system;
 KW Alzheimer's disease; tumour; high affinity; receptor antagonist;
 KW development; non-steroidal; anti-inflammatory.

XX
 OS Homo sapiens.

XX
 FH Key Location/Qualifiers
 FT Region 4..7
 FT /note= "beta-turn"
 FT Misc-difference 9 /note= "D-form residue"
 FT Modified-site 7 /note= "N-methylated glutamine"

XX
 PN W09606629-A1.
 XX
 PD 07-MAR-1996.
 XX
 PF 31-AUG-1995: 95WO-US11126.
 XX
 PF 31-AUG-1994: 94US-0299285.
 XX
 PR (UYNE-) UNIV NEBRASKA.
 XX (UYOU) UNIV QUEENSLAND.
 PA
 PA (UYOU) UNIV QUEENSLAND.
 XX
 PI Klnarsky L, Sanderson DG, Sherman SA, Taylor SM;
 DR WPI: 1996-160140/16.
 XX
 XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
 PT spasmogenesis, platelet aggregation and increases in cell membrane
 PT permeability
 XX
 PS Example 5: Page 91: 116pp: English.
 XX
 CC The present peptide is an analogue of a C-terminal decapeptide
 CC region from human C5a anaphylatoxin, which has a constrained

CC backbone conformation (compared to the natural peptide's flexible
 CC C-terminal region) comprising a beta-turn. Analogues with type
 CC (II or V) and type III beta-turns elicit proinflammatory responses
 CC characterised by spasmogenesis, platelet aggregation and neutrophil
 CC non-mediated increases of cell membrane (specifically vascular)
 CC permeability, or neutrophil polarisation, neutrophil enzyme
 CC release and neutrophil mediated increases in cell membrane (esp.
 CC vascular) permeability, respectively.

CC The analogues can be used as immune adjuvants for the treatment of
 CC immune deficiency disorders, and for augmenting standard immune
 CC therapy for the treatment of cancer, without inflammatory side
 CC effects. They can also be used to facilitate drug delivery by
 CC increasing vascular permeability, e.g. in the treatment of CNS
 CC disorders such as Alzheimer's disease or tumours, and to develop
 CC high affinity C5a receptor antagonists useful as non-steroidal
 CC anti-inflammatory agents.

XX
 CC Sequence 10 AA:

SO

Query Match 88.7%: Score 47; DB 17; Length 10;
 Best Local Similarity 90.0%; Pred. No. 0.0024;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSFKDMQLGR 10
 |||||
 Db 1 ysfkdmqlar 10

RESULT 8
 AAR94495
 ID AAR94495 standard; peptide: 10 AA.
 AC AAR94495;
 XX
 DT 17-OCT-1996 (first entry)

C5a anaphylatoxin analogue C5a(65-74) Tyr65 NME-D-Ala73 NME-Arg74.

XX
 KW C-terminal: human; C5a anaphylatoxin; analogue: beta-turn;
 KW constrained backbone conformation: proinflammatory response;
 KW spasmogenesis: platelet aggregation; neutrophil: non-mediated;
 KW mediated; increase; cell membrane; vascular; permeability;
 KW neutrophil polarisation; neutrophil enzyme release; treatment;
 KW immune adjuvant; immunodeficiency; augmentation; immune therapy;
 KW cancer; drug delivery; CNS disorder; central nervous system;
 KW Alzheimer's disease; tumour; high affinity; receptor antagonist;
 KW development; non-steroidal; anti-inflammatory.

XX
 CS Homo sapiens.

XX
 FH Key Location/Qualifiers
 FT Region 4..7
 FT /note= "beta-turn"
 FT Modified-site 9 /note= "N-methylated D-alanine"
 FT Modified-site 10 /note= "N-methylated arganine"

XX
 PN W09606629-A1.
 XX
 PD 07-MAR-1996.
 XX
 PF 31-AUG-1995: 95WO-US11126.
 XX
 PF 31-AUG-1994: 94US-0299285.
 XX
 PR (UYNE-) UNIV NEBRASKA.
 XX (UYOU) UNIV QUEENSLAND.
 PA
 PA (UYOU) UNIV QUEENSLAND.
 XX
 PI Klnarsky L, Sanderson DG, Sherman SA, Taylor SM;
 DR WPI: 1996-160140/16.

```

XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
PT spasmogenesis; platelet aggregation and increases in cell membrane
PT permeability
PS
XX Example 5: Page 92; 116pp: English.
XX
CC The present peptide is an analogue of a C-terminal decapeptide
CC region from human C5a anaphylatoxin, which has a constrained
CC backbone conformation (compared to the natural peptide's flexible
CC C-terminal region) comprising a beta-turn. Analogues with type
CC (II or V) and type III beta-turns elicit proinflammatory responses
CC characterised by spasmogenesis; platelet aggregation and neutrophil
CC non-mediated increases of cell membrane (specifically vascular)
CC permeability, or neutrophil polarisation, neutrophil enzyme
CC release and neutrophil mediated increases in cell membrane (esp.
CC vascular) permeability, respectively.
CC The analogues can be used as immune adjuvants for the treatment of
CC immune deficiency disorders, and for augmenting standard immune
CC therapy for the treatment of cancer, without inflammatory side
CC effects. They can also be used to facilitate drug delivery by
CC increasing vascular permeability, e.g. in the treatment of CNS
CC disorders such as Alzheimer's disease or tumours, and to develop
CC high affinity C5a receptor antagonists useful as non-steroidal
CC anti-inflammatory agents.
XX
SQ Sequence 10 AA:

```

```

Query Match      88.7%: Score 47; DB 17; Length 10;
Best Local Similarity 90.0%: Pred. No. 0.0024;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 YSFKDMQLCR 10
   | | | | | | | |
DB 1 ysfkdmqlar 10

```

```

RESULT 9
AAR94496
ID AAR94496 standard; peptide: 10 AA.
XX
AC AAR94496:
XX
DT 17-OCT-1996 (first entry)
XX
DE C5a analogue C5a(65-74) Tyr65 NMe-Leu72 D-Ala73 NMe-Arg74.
XX
KW C-terminal; human; C5a anaphylatoxin; analogue; beta-turn;
KW constrained backbone conformation; proinflammatory response;
KW spasmogenesis; platelet aggregation; neutrophil; non-mediated;
KW mediated; increase; cell membrane; vascular; permeability;
KW neutrophil polarisation; neutrophil enzyme release; treatment;
KW immune adjuvant; immunodeficiency; augmentation; immune therapy;
KW cancer; drug delivery; CNS disorder; central nervous system;
KW Alzheimer's disease; tumour; high affinity; receptor antagonist;
KW development; non-steroidal; anti-inflammatory.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Modified-site 8 /note="N-methylated leucine"
FT Misc-difference 9 /note="D-form residue"
FT Modified-site 10 /note="N-methylated arginine"
XX
XX W09606629-A1.
XX
XX 07-MAR-1996.
XX
XX 31-AUG-1995; 95MO-US11126.

```

```

XX 31-AUG-1994; 94US-0299285.
PR (UYNE-) UNIV NEBRASKA.
PA (UYOU ) UNIV QUEENSLAND.
XX
XX Kinnarsky L., Sanderson DG, Sherman SA, Taylor SM:
XX WPI: 1996-160140/16.
XX
XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
PT spasmogenesis; platelet aggregation and increases in cell membrane
PT permeability
PS
XX Example 5: Page 93; 116pp: English.
XX
CC The present peptide is an analogue of a C-terminal decapeptide
CC region from human C5a anaphylatoxin, which has a constrained
CC backbone conformation (compared to the natural peptide's flexible
CC C-terminal region) comprising a beta-turn. Analogues with type
CC (II or V) and type III beta-turns elicit proinflammatory responses
CC characterised by spasmogenesis; platelet aggregation and neutrophil
CC non-mediated increases of cell membrane (specifically vascular)
CC permeability, or neutrophil polarisation, neutrophil enzyme
CC release and neutrophil mediated increases in cell membrane (esp.
CC vascular) permeability, respectively.
CC The analogues can be used as immune adjuvants for the treatment of
CC immune deficiency disorders, and for augmenting standard immune
CC therapy for the treatment of cancer, without inflammatory side
CC effects. They can also be used to facilitate drug delivery by
CC increasing vascular permeability, e.g. in the treatment of CNS
CC disorders such as Alzheimer's disease or tumours, and to develop
CC high affinity C5a receptor antagonists useful as non-steroidal
CC anti-inflammatory agents.
XX
SQ Sequence 10 AA:

```

```

Query Match      88.7%: Score 47; DB 17; Length 10;
Best Local Similarity 90.0%: Pred. No. 0.0024;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 YSFKDMQLCR 10
   | | | | | | | |
DB 1 ysfkdmqlar 10

```

```

RESULT 10
AAR94497
ID AAR94497 standard; peptide: 10 AA.
XX
AC AAR94497:
XX
DT 17-OCT-1996 (first entry)
XX
DE C5a analogue C5a(65-74) Tyr65 NMe-Gln71 D-Ala73 NMe-Arg74.
XX
KW C-terminal; human; C5a anaphylatoxin; analogue; beta-turn;
KW constrained backbone conformation; proinflammatory response;
KW spasmogenesis; platelet aggregation; neutrophil; non-mediated;
KW mediated; increase; cell membrane; vascular; permeability;
KW neutrophil polarisation; neutrophil enzyme release; treatment;
KW immune adjuvant; immunodeficiency; augmentation; immune therapy;
KW cancer; drug delivery; CNS disorder; central nervous system;
KW Alzheimer's disease; tumour; high affinity; receptor antagonist;
KW development; non-steroidal; anti-inflammatory.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Modified-site 7 /note="N-methylated glutamine"
FT Misc-difference 9

```

FT Modified-site /note="D-form residue"
 FT 10
 FT /note="N-methylated arginine"
 XX
 XX
 PN WO9606629-A1.
 XX
 PD 07-MAR-1996.
 XX
 PF 31-AUG-1995; 95WO-US11126.
 XX
 PR 31-AUG-1994; 94US-0299285.
 XX
 PA (UYNE-) UNIV NEBRASKA.
 PA (UYOU) UNIV QUEENSLAND.
 XX
 PI Kirnarsky L, Sanderson DG, Sherman SA, Taylor SM:
 DR WPI; 1996-160140/16.
 XX
 XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
 PT spasmogenesis, platelet aggregation and increases in cell membrane
 PT permeability
 XX
 PS Example 5; Page 93; 116pp: English.
 XX
 CC The present peptide is an analogue of a C-terminal decapeptide
 CC region from human C5a anaphylatoxin, which has a constrained
 CC backbone conformation (compared to the natural peptide's flexible
 CC C-terminal region) comprising a beta-turn. Analogues with type
 CC (II or V) and type III beta-turns elicit proinflammatory responses
 CC characterised by spasmogenesis, platelet aggregation and neutrophil
 CC non-mediated increases of cell membrane (specifically vascular)
 CC permeability, or neutrophil polarisation, neutrophil enzyme
 CC release and neutrophil mediated increases in cell membrane (esp.
 CC vascular) permeability, respectively.
 CC The analogues can be used as immune adjuvants for the treatment of
 CC immune deficiency disorders, and for augmenting standard immune
 CC therapy for the treatment of cancer, without inflammatory side
 CC effects. They can also be used to facilitate drug delivery by
 CC increasing vascular permeability, e.g. in the treatment of CNS
 CC disorders such as Alzheimer's disease or tumours, and to develop
 CC high affinity C5a receptor antagonists useful as non-steroidal
 CC anti-inflammatory agents.
 CC
 XX
 SQ Sequence 10 AA:
 OY 1 YSFKDQGLR 10
 DB 1 YSFKDQGLR 10
 XX
 XX
 XX
 DE C5a analogue C5a(65-74) Tyr65 NME-Gln71 D-Ala73 NME-Leu72.
 XX
 XX C-terminal; human; C5a anaphylatoxin; analogue; beta-turn;
 KM constrained backbone conformation; proinflammatory response;
 KM spasmogenesis; platelet aggregation; neutrophil; non-mediated;
 KM mediated; increase; cell membrane; vascular; permeability;
 KM neutrophil polarisation; neutrophil enzyme release; treatment;
 KM immune adjuvant; immunodeficiency; augmentation; immune therapy;
 KM cancer; drug delivery; CNS disorder; central nervous system;

KW Alzheimer's disease; tumour; high affinity; receptor antagonist;
 KW development; non-steroidal; anti-inflammatory.
 XX
 XX Homo sapiens.
 CS
 CS
 FH Key Location/Qualifiers
 FT Modified-site 7
 FT Misc-difference 9 /note="N-methylated glutamine"
 FT /note="D-form residue"
 FT Modified-site 8 /note="N-methylated leucine"
 FT /note="N-methylated arginine"
 XX
 PN WO9606629-A1.
 XX
 PD 07-MAR-1996.
 XX
 PF 31-AUG-1995; 95WO-US11126.
 XX
 PR 31-AUG-1994; 94US-0299285.
 XX
 PA (UYNE-) UNIV NEBRASKA.
 PA (UYOU) UNIV QUEENSLAND.
 XX
 PI Kirnarsky L, Sanderson DG, Sherman SA, Taylor SM:
 DR WPI; 1996-160140/16.
 XX
 XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
 PT spasmogenesis, platelet aggregation and increases in cell membrane
 PT permeability
 XX
 PS Example 5; Page 94; 116pp: English.
 XX
 CC The present peptide is an analogue of a C-terminal decapeptide
 CC region from human C5a anaphylatoxin, which has a constrained
 CC backbone conformation (compared to the natural peptide's flexible
 CC C-terminal region) comprising a beta-turn. Analogues with type
 CC (II or V) and type III beta-turns elicit proinflammatory responses
 CC characterised by spasmogenesis, platelet aggregation and neutrophil
 CC non-mediated increases of cell membrane (specifically vascular)
 CC permeability, or neutrophil polarisation, neutrophil enzyme
 CC release and neutrophil mediated increases in cell membrane (esp.
 CC vascular) permeability, respectively.
 CC The analogues can be used as immune adjuvants for the treatment of
 CC immune deficiency disorders, and for augmenting standard immune
 CC therapy for the treatment of cancer, without inflammatory side
 CC effects. They can also be used to facilitate drug delivery by
 CC increasing vascular permeability, e.g. in the treatment of CNS
 CC disorders such as Alzheimer's disease or tumours, and to develop
 CC high affinity C5a receptor antagonists useful as non-steroidal
 CC anti-inflammatory agents.
 CC
 XX
 SQ Sequence 10 AA:
 OY 1 YSFKDQGLR 10
 DB 1 YSFKDQGLR 10
 XX
 XX
 XX
 DE C-terminal; human; C5a anaphylatoxin; analogue; beta-turn;
 KM constrained backbone conformation; proinflammatory response;
 KM spasmogenesis; platelet aggregation; neutrophil; non-mediated;
 KM mediated; increase; cell membrane; vascular; permeability;
 KM neutrophil polarisation; neutrophil enzyme release; treatment;
 KM immune adjuvant; immunodeficiency; augmentation; immune therapy;
 KM cancer; drug delivery; CNS disorder; central nervous system;

DE C5a analogue C5a(65-74) Tyr65 NME-D-Ala73 NME-Leu72.

XX C-terminal; human; C5a anaphylatoxin; analogue; beta-turn;
 KW constrained backbone conformation; proinflammatory response;
 KW spasmogenesis; platelet aggregation; neutrophil; non-mediated;
 KW mediated; increase; cell membrane; vascular; permeability;
 KW neutrophil polarisation; neutrophil enzyme release; treatment;
 KW immune adjuvant; immunodeficiency; augmentation; immune therapy;
 KW cancer; drug delivery; CNS disorder; central nervous system;
 KW Alzheimer's disease; tumour; high affinity; receptor antagonist;
 KW development; non-steroidal; anti-inflammatory.

OS Homo sapiens.

XX Key Location/Qualifiers
 FH Modified-site 9 /note= "N-methylated D-alanine"
 FT Modified-site 8 /note= "N-methylated leucine"
 FT

PN WO9606629-A1.

XX 07-MAR-1996.

XX 31-AUG-1995; 95WO-US11126.

XX 31-AUG-1994; 94US-0299285.

XX (UYNE-) UNIV NEBRASKA.

PA (UYOU) UNIV QUEENSLAND.

XX Kirarsky L, Sanderson DG, Sherman SA, Taylor SM;

PI WPI; 1996-160140/16.

XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
 PT spasmogenesis, platelet aggregation and increases in cell membrane
 PT permeability

XX Example 5; Page 94; 116pp; English.

XX The present peptide is an analogue of a C-terminal decapeptide
 CC region from human C5a anaphylatoxin, which has a constrained
 CC backbone conformation (compared to the natural peptide's flexible
 CC C-terminal region) comprising a beta-turn. Analogues with type
 CC (II or V) and type III beta-turns elicit proinflammatory responses
 CC characterised by spasmogenesis, platelet aggregation and neutrophil
 CC non-mediated increases of cell membrane (specifically vascular)
 CC permeability, or neutrophil polarisation, neutrophil enzyme
 CC release and neutrophil mediated increases in cell membrane (esp.
 CC vascular) permeability, respectively.
 CC The analogues can be used as immune adjuvants for the treatment of
 CC immune deficiency disorders, and for augmenting standard immune
 CC effects. They can also be used to facilitate drug delivery by
 CC increasing vascular permeability, e.g. in the treatment of CNS
 CC disorders such as Alzheimer's disease or tumours, and to develop
 CC high affinity C5a receptor antagonists useful as non-steroidal
 CC anti-inflammatory agents.

XX Sequence 10 AA:

Query Match

Best Local Similarity 88.7%; Score 47; DB 17; Length 10;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 YSFKDWOLCR 10

DB 1 Ysfkdwqlar 10

RESULT 13

AAR94500
 ID AAR94500 standard; peptide; 10 AA.

XX AAR94500;

XX 17-OCT-1996 (first entry)

XX C5a analogue C5a(65-74) Tyr65 NME-D-Ala73 NME-Leu72 NME-Arg74.

XX C-terminal; human; C5a anaphylatoxin; analogue; beta-turn;
 KW constrained backbone conformation; proinflammatory response;
 KW spasmogenesis; platelet aggregation; neutrophil; non-mediated;
 KW mediated; increase; cell membrane; vascular; permeability;
 KW neutrophil polarisation; neutrophil enzyme release; treatment;
 KW immune adjuvant; immunodeficiency; augmentation; immune therapy;
 KW cancer; drug delivery; CNS disorder; central nervous system;
 KW Alzheimer's disease; tumour; high affinity; receptor antagonist;
 KW development; non-steroidal; anti-inflammatory.

OS Homo sapiens.

XX Key Location/Qualifiers
 FH Modified-site 10 /note= "N-methylated arginine"
 FT Modified-site 9 /note= "N-methylated D-alanine"
 FT Modified-site 8 /note= "N-methylated leucine"
 FT

PN WO9606629-A1.

XX 07-MAR-1996.

XX 31-AUG-1995; 95WO-US11126.

XX 31-AUG-1994; 94US-0299285.

XX (UYNE-) UNIV NEBRASKA.

PA (UYOU) UNIV QUEENSLAND.

XX Kirarsky L, Sanderson DG, Sherman SA, Taylor SM;

PI WPI; 1996-160140/16.

XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
 PT spasmogenesis, platelet aggregation and increases in cell membrane
 PT permeability

XX Example 5; Page 95; 116pp; English.

XX The present peptide is an analogue of a C-terminal decapeptide
 CC region from human C5a anaphylatoxin, which has a constrained
 CC backbone conformation (compared to the natural peptide's flexible
 CC C-terminal region) comprising a beta-turn. Analogues with type
 CC (II or V) and type III beta-turns elicit proinflammatory responses
 CC characterised by spasmogenesis, platelet aggregation and neutrophil
 CC non-mediated increases of cell membrane (specifically vascular)
 CC permeability, or neutrophil polarisation, neutrophil enzyme
 CC release and neutrophil mediated increases in cell membrane (esp.
 CC vascular) permeability, respectively.
 CC The analogues can be used as immune adjuvants for the treatment of
 CC immune deficiency disorders, and for augmenting standard immune
 CC effects. They can also be used to facilitate drug delivery by
 CC increasing vascular permeability, e.g. in the treatment of CNS
 CC disorders such as Alzheimer's disease or tumours, and to develop
 CC high affinity C5a receptor antagonists useful as non-steroidal
 CC anti-inflammatory agents.

XX Sequence 10 AA:

Query Match

88.7%; Score 47; DB 17; Length 10;

Best Local Similarity: 90.0%; Pred. No. 0.0024;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 YSFKMOGLR 10
 DB 1 ysfkmoqlar 10

RESULT 14

AAR94471

ID AAR94471 standard: peptide: 10 AA.

XX AAR94471;

XX 17-OCT-1996 (first entry)

XX C5a anaphylatoxin C-terminal analogue C5a(65-74) Tyr65 D-Ala73.

XX C-terminal; human; C5a anaphylatoxin; analogue; beta-turn;

XX constrained backbone conformation; proinflammatory response;

XX spasmogenesis; platelet aggregation; neutrophil; non-mediated;

XX mediated; increase; cell membrane; vascular; permeability;

XX neutrophil polarisation; neutrophil enzyme release; treatment;

XX immune adjuvant; immunodeficiency; augmentation; immune therapy;

XX cancer; drug delivery; CNS disorder; central nervous system;

XX Alzheimer's disease; tumour; high affinity; receptor antagonist;

XX development; non-steroidal; anti-inflammatory.

XX Homo sapiens.

XX Key

XX Region 4..7

XX MISC-difference 9 /note= "beta-turn"

XX MISC-difference 9 /note= "D-form residue"

XX MO9606629-A1.

XX 07-MAR-1996.

XX 31-AUG-1995: 95WO-US11126.

XX 31-AUG-1994: 94US-0299285.

XX (UYNE-) UNIV NEBRASKA.

XX (UYOU) UNIV QUEENSLAND.

XX Kirnarsky L, Sanderson DG, Sherman SA, Taylor SM;

XX MPI: 1996-160140/16.

XX C-terminal analogues of C5a anaphylatoxin - induce e.g.

XX spasmogenesis, platelet aggregation and increases in cell membrane

XX permeability

XX Claim 8: Page 80: 116pp: English.

XX The present peptide is an analogue of a C-terminal decapeptide

XX region from human C5a anaphylatoxin, which has a constrained

XX backbone conformation (compared to the natural peptide's flexible

XX C-terminal region) comprising a beta-turn. Analogues with type

XX (II or V) and type III beta-turns elicit proinflammatory responses

XX characterised by spasmogenesis, platelet aggregation and neutrophil

XX non-mediated increases of cell membrane (specifically vascular)

XX permeability, or neutrophil polarisation, neutrophil enzyme

XX release and neutrophil mediated increases in cell membrane (esp.

XX vascular) permeability, respectively.

XX The analogues can be used as immune adjuvants for the treatment of

XX immune deficiency disorders, and for augmenting standard immune

XX therapy for the treatment of cancer, without inflammatory side

XX effects. They can also be used to facilitate drug delivery by

XX increasing vascular permeability, e.g. in the treatment of CNS

XX disorders such as Alzheimer's disease or tumours, and to develop

XX

XX

XX

XX

XX

XX

XX

high affinity C5a receptor antagonists useful as non-steroidal
 anti-inflammatory agents.
 In a type (II or V) beta-turn specific human foetal artery smooth
 muscle contraction assay, and a type III beta-turn specific
 neutrophil polarisation assay the peptide had respective EC50
 (microm) activities of 12.2 and 1.2, compared to 0.018 and 0.0013
 for C5a.

Sequence 10 AA:

Query Match 88.7%; Score 47; DB 17; Length 10;

Best Local Similarity 90.0%; Pred. No. 0.0024;

Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSFKMOGLR 10

DB 1 ysfkmoqlar 10

RESULT 15

AAR94468

ID AAR94468 standard: peptide: 10 AA.

XX AAR94468;

XX 17-OCT-1996 (first entry)

XX C5a anaphylatoxin C-terminal analogue C5a(65-74) Tyr65 Pro71.

XX C-terminal; human; C5a anaphylatoxin; analogue; beta-turn;

XX constrained backbone conformation; proinflammatory response;

XX spasmogenesis; platelet aggregation; neutrophil; non-mediated;

XX mediated; increase; cell membrane; vascular; permeability;

XX neutrophil polarisation; neutrophil enzyme release; treatment;

XX immune adjuvant; immunodeficiency; augmentation; immune therapy;

XX cancer; drug delivery; CNS disorder; central nervous system;

XX Alzheimer's disease; tumour; high affinity; receptor antagonist;

XX development; non-steroidal; anti-inflammatory.

XX Homo sapiens.

XX Key

XX Region 4..7

XX MISC-difference 9 /note= "beta-turn"

XX MO9606629-A1.

XX 07-MAR-1996.

XX 31-AUG-1995: 95WO-US11126.

XX 31-AUG-1994: 94US-0299285.

XX (UYNE-) UNIV NEBRASKA.

XX (UYOU) UNIV QUEENSLAND.

XX Kirnarsky L, Sanderson DG, Sherman SA, Taylor SM;

XX MPI: 1996-160140/16.

XX C-terminal analogues of C5a anaphylatoxin - induce e.g.

XX spasmogenesis, platelet aggregation and increases in cell membrane

XX permeability

XX Claim 8: Page 78: 116pp: English.

XX The present peptide is an analogue of a C-terminal decapeptide

XX region from human C5a anaphylatoxin, which has a constrained

XX backbone conformation (compared to the natural peptide's flexible

XX C-terminal region) comprising a beta-turn. Analogues with type

XX (II or V) and type III beta-turns elicit proinflammatory responses

XX characterised by spasmogenesis, platelet aggregation and neutrophil

XX

XX

XX

XX

XX

XX

XX

XX

CC non-mediated increases of cell membrane (specifically vascular)
CC permeability, or neutrophil polarisation, neutrophil enzyme
CC release and neutrophil mediated increases in cell membrane (esp.
CC vascular) permeability, respectively.
CC The analogues can be used as immune adjuvants for the treatment of
CC immune deficiency disorders, and for augmenting standard immune
CC therapy for the treatment of cancer, without inflammatory side
CC effects. They can also be used to facilitate drug delivery by
CC increasing vascular permeability, e.g. in the treatment of CNS
CC disorders such as Alzheimer's disease or tumours, and to develop
CC high affinity C5a receptor antagonists useful as non-steroidal
CC anti-inflammatory agents.
CC In a type (II or V) beta-turn specific human foetal artery smooth
CC muscle contraction assay, and a type III beta-turn specific
CC neutrophil polarisation assay the peptide had respective EC50
CC (microm) activities of 1.13 and 5, compared to 0.018 and 0.0013
CC for C5a.
CC
XX

SQ Sequence 10 AA:

Query Match

Best Local Similarity 88.7%; Score 47; DB 17; Length 10;
Pred. No. 0.0024;

Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 YSFKDMQLGR 10
| | | | | | | | | |
DB 1 ysfkdmplgr 10

Search completed: February 27, 2002, 11:41:13
Job time: 452 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:45:10 ; Search time 145.23 seconds
(without alignments)
5.245 Million cell updates/sec

Title: US-09-446-109a-2
Perfect score: 53
Sequence: 1 YSFKDMQLCR 10

Scoring table: BLOSUM62
Gapop 10.0, Capext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	39	73.6	571 2 DB1334	hydrogenase (EC 1.1.1.1)
2	39	73.6	1676 1 CSHU	complement C5 prec
3	36	67.9	791 2 H96839	hypothetical prote
4	35	66.0	283 2 T19411	hypothetical prote
5	35	66.0	833 2 T22139	hypothetical prote
6	34	64.2	319 1 OMBS	ornithine carbamoy
7	34	64.2	407 2 DS3308	mosc protein - Rhl
8	34	64.2	574 2 R44552	similar to Mlo pro
9	34	64.2	661 1 OXRTAI	acyl-CoA oxidase (
10	34	64.2	661 1 OXRTA2	acyl-CoA oxidase (
11	34	64.2	797 2 H83190	probable outer mem
12	34	64.2	1413 2 G84790	probable ABC trans
13	33	62.3	203 2 DB1328	probable riboflavi
14	33	62.3	216 2 T02418	hypothetical prote
15	33	62.3	225 2 G75167	hypothetical prote
16	33	62.3	236 2 G71047	hypothetical prote
17	33	62.3	242 2 T17575	procollagen-prolin
18	33	62.3	305 1 PWRVNB	Na+/K+-exchanging
19	33	62.3	349 2 G83947	multifunctional pr
20	33	62.3	385 2 D12858	late expression fa
21	33	62.3	385 2 T18111	LEF-3 orf67 - Bomb
22	33	62.3	388 2 S30219	recombination prot
23	33	62.3	438 2 A46259	recombination prot
24	33	62.3	439 2 T01050	hypothetical prote
25	33	62.3	541 1 H64163	probable 60k inner
26	33	62.3	678 2 S48412	probable membrane
27	33	62.3	1738 2 S20614	conserved hypothet
28	32	60.4	115 2 G81438	probable periplasm
29	32	60.4	170 2 D69887	conserved hypothet

30	32	60.4	217 1 G70459	conserved hypothet
31	32	60.4	239 2 E71911	ribonuclease III -
32	32	60.4	240 2 F64602	transposase III -
33	32	60.4	241 2 G84149	transposase (26) B
34	32	60.4	266 1 ENSAB6	enterotoxin B prec
35	32	60.4	320 2 T47882	hypothetical prote
36	32	60.4	343 2 F81264	recombination prot
37	32	60.4	343 2 F81264	reca protein Cj167
38	32	60.4	356 2 H84934	UDP-N-acetylmurama
39	32	60.4	460 2 G96764	unknown protein F2
40	32	60.4	468 2 B69667	Na+/H+-exchanging
41	32	60.4	554 2 S41526	hemolysin accessor
42	32	60.4	592 2 B82498	sulfate permease f
43	32	60.4	726 2 T31287	hypothetical prote
44	32	60.4	930 2 C02083	DNA-PK - human (fr
45	32	60.4	2088 2 E71436	hypothetical prote

ALIGNMENTS

RESULT 1
DB1334
hydrogenase (EC 1.18.99.1) (Nife) large chain Cj1266c [similarity] - Campylobacter je
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 20-Jun-2000
C:Accession: DB1334
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Basham, D.; Chit
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; Vaviliet, A.; Whitehead, S.; Ba
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A:Reference number: AB1250; MUID:20150912
A:Accession: DB1334
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-571 <PAR>
A:Cross-references: GB:AL139077; GB:AL111168; NID:6968444; PID:CA873520.L; PID:696
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: hydB; Cj1266c
C:Superfamily: hydrogenase (Nife) large chain
C:Keywords: iron; metalloprotein; nickel; oxidoreductase
F:62,65,547,550/Binding site: nickel (Cys) #status predicted
F:65,550/Binding site: iron (Cys) #status predicted
F:69/Active site: His #status predicted

Query Match 73.6% Score 39; DB 2; Length 571;
Best Local Similarity 60.0%; Pred. No. 4.2;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSFKDMQLCR 10
DB 282 YTFKFOICR 291

RESULT 2
CSHU
complement C5 precursor (validated) - human
N:Contains: C5a anaphylatoxin; C5b
C:Species: Homo sapiens (man)
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 08-Dec-2000
C:Accession: A40075; A27689; A01267; A01266; S15121
R:Harland, D.L.; Harland, J.C.; Fleischer, D.T.; Hunt, A.; Wetsel, R.A.
J. Immunol. 146, 362-368, 1991
A:Title: Complete cDNA sequence of human complement pro-C5. Evidence of truncated tra
A:Reference number: A40075; MUID:91079575
A:Accession: A40075
A:Molecule type: mRNA
A:Residues: 1-1676 <NAV>
A:Cross-references: GB:M57729; NID:9179982; PID:AAA51925.1; PID:9179983
Note: 518-Ser was also found
R:Wetsel, R.A.; Lemons, K.S.; Le Beau, M.M.; Barnum, S.R.; Noack, D.; Tack, B.F.

Biochemistry 27 1474-1482, 1988

A>Title: Molecular analysis of human complement component C5: localization of the structural gene.

A:Reference number: A27609; MUID:8820511

A:Molecule type: mRNA

A:Residues: 412-1676 <MET>

A:Cross-references: GB:065134; GB:ML8879; NID:g179691; PUDN:AAAS1856.L; PID:g179692

J.Biol.Chem. 253, 6955-6964, 1978

R.Fernandez, H.N.; Hugli, T.E.

A>Title: Primary structural analysis of the polypeptide portion of human C5a anaphylatoxin.

A:Reference number: A01267; MUID:79005687

A:Accession: A01267

A:Molecule type: protein

A:Residues: 678-751 <FER>

R.Lundvall, A.B.; Welser, R.A.; Kristensen, T.; Whitehead, A.S.; Woods, D.E.; Ogden, R.C.

J. Biol. Chem. 260, 2108-2112, 1985

A>Title: Isolation and sequence analysis of a cDNA clone encoding the fifth complement component.

A:Reference number: A01266; MUID:85130937

A:Accession: A01266

A:Molecule type: mRNA

A:Residues: 412-834, SLALSPLECNKISGHCKRLRPGSSDPSASQVAGITNNHADR' <LUN>

A:Cross-references: GB:K02874

A>Note: The carboxyl-terminal part of the sequence in this report appears to be derived from the same gene as the one reported by Bohnsack et al. (1985).

R.Bohnsack, J.F.; Mollison, K.W.; Buko, A.M.; Ashworth, J.C.; Hill, H.R.

Biochem. J. 273, 635-640, 1991

A>Title: Group B streptococci inactivate complement component C5a by enzymic cleavage at the C5a-C5b junction.

A:Reference number: S15121; MUID:91144547

A:Contents: annotation

C:Comment: complement C5 contains two disulfide-linked chains, formed by removal of four amino acids from the C5 protein.

(beta and alpha' chains).

C:Comment: Activation of C5 initiates the spontaneous assembly of the late complement components C5b, C6, C7, C8, and C9.

Is the foundation upon which the membrane attack complex is assembled.

C:Comment: C5a has potent spasmogenic and chemotactic activity.

C:Genetics:

A:Gene: GDB:C5

A:Cross-references: GDB:119734; OMIM:120900

A:Map position: 9q34.1-q34.1

A:Superfamily: alpha-2-macroglobulin

C:Keywords: complement alternate pathway; complement pathway; cytolytic; glycoprotein; immunology; protein structure; signal sequence

F:1-18/Domains: signal sequence #status predicted <SIG>

F:1-18/Products: complement C5 #status predicted <C5T>

F:19-673,678-1676/Product: complement C5 #status predicted <C5B>

F:19-673,752-1676/Product: C5b #status predicted <C5B>

F:19-673/Product: complement C5 and C5b beta chain #status predicted <C5BB>

F:678-1676/Product: complement C5 and C5b alpha chain #status predicted <C5AB>

F:678-751/Product: C5a anaphylatoxin #status experimental <C5A>

F:752-1676/Product: C5b alpha' chain #status predicted <C5BA>

F:752-810,634-659,698-731,711-732,866-1527,1101-1159,1375-1505,1405-1474,1520-1515/Binding site: carbohydrate (Asn) (covalent) #status experimental

F:751-752/Cleavage site: carbohydrate (Asn) (covalent) #status experimental

F:751,1115,1630/Binding site: carbohydrate (Asn) (covalent) #status predicted

```

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallio
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: H96839
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-791 <STO>
A:Cross-references: GB:AE005173; NID:96503286; PIDN:AAF14662.1; GSPDB:GN00141
A:Genetics:
A:Gene: F23A5.9
A:Map position: 1

Query Match      67.9%; Score 36; DB 2; Length 791;
Best Local Similarity 75.0%; Pred. No. 25;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Caps 0;

QY      3 FKDMQLR 10
      11111111
Db      537 FKDLQVCR 544

RESULT      4
T19411
hypothetical protein C23H4.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_rev15on 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T19411
R:Wilkinson, J
submitted to the EMBL Data Library, August 1996
A:Reference number: Z19121
A:Accession: T19411
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-283 <MTL>
A:Cross-references: EMBL:Z78416; PIDN:CAE01678.1; GSPDB:GN00028; CESP:C23H4.4
A:Experimental source: clone C23H4
C:Genetics:
A:Gene: CESP:C23H4.4
A:Map position: X
A:Introns: 64/3; 97/2; 130/3; 198/1

Query Match      66.0%; Score 35; DB 2; Length 283;
Best Local Similarity 55.6%; Pred. No. 13;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      1 YSFKMDQLG 9
      11111111
Db      260 YGFEDLQIC 268

RESULT      5
T22139
hypothetical protein F43G6.9 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_rev15on 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T22139; T22174
R:Swindburne, J.
submitted to the EMBL Data Library, July 1995
A:Reference number: Z19522
A:Accession: T22139
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-833 <MTL>
A:Cross-references: EMBL:Z50070; NID:el519046; PIDN:CAA90402.1; GSPDB:GN00020; CESP:F
R:Matthews, L.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19526
A:Accession: T22174

```


A:Status: preliminary: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-833 <M12>
 A:Cross-references: EMBL:Z83108; P1DN:CA805512.1; GSPDB:GM00020; CESP:F43G6.9
 A:Experimental source: clone F44E5
 C:Genetics:
 A:Gene: CESP:F43G6.9
 A:Map position: 2
 A:Introns: 21/1: 68/3; 122/3; 186/3; 381/3; 474/3; 641/2; 752/3

Query Match 66.0%; Score 35; DB 2; Length 833;
 Best Local Similarity 75.0%; Pred. No. 43;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFKDMQLG 9
 Db 205 AFKDLQLG 212

RESULT 6
 ornithine carbamoyltransferase (EC 2.1.3.3) - Bacillus subtilis
 N:Alternate names: citrulline phosphorilase; ornithine transcarbamylase
 C:Species: Bacillus subtilis
 C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Jun-2000
 C:Accession: S11000; A38768; I40378; A69589; S38434
 R:Mountain, A.; Smith, M.C.M.; Baumberg, S.
 Nucleic Acids Res. 18, 4594, 1990
 A:Title: Nucleotide sequence of the Bacillus subtilis argF gene encoding ornithine carba
 A:Reference number: S10999; MUID:90356402
 A:Accession: S11000
 A:Molecule type: DNA
 A:Residues: 1-319 <M0U>
 A:Cross-references: EMBL:X53360; NID:q39809; P1DN:CAA37444.1; P1D:q39811
 A:Accession: A38768
 A:Molecule type: protein
 A:Residues: 1,'X','3-14,'XX',17-20,'X',22-30,'XX',33,'X',35-39 <M0U2>
 R:O'Reilly, M.; Devine, K.M.
 Microbiology 140, 1023-1025, 1994
 A:Title: Sequence and analysis of the citrulline biosynthetic operon argC-F from Bacillu
 A:Reference number: I40372; MUID:94297722
 A:Accession: I40378
 A:Status: preliminary: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-319 <RES>
 A:Cross-references: EMBL:Z26919; NID:q408113; P1DN:CA81542.1; P1D:q408120
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berche
 C.: Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
 A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gall
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hult, M.F.
 Koeller, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardin
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
 Y. M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portelle
 Rieger, M.; Rivolta, C.; Rooha, E.; Roche, B.; Rose, M.; Sadleir, J.; Sato, T.; Scanlon
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot
 Aheuchl, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; MUID:98044033
 A:Accession: A69589
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-319 <KUN>
 A:Cross-references: GB:Z99110; GB:AL009126; GB:AL009126; NID:92633472; P1DN:CA812982.1; P1
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: argF
 A:Map position: 100 (degrees)
 C:Superfamily: ornithine carbamoyltransferase; aspartate/ornithine carbamoyltransferase

C:Keywords: arginine biosynthesis; transferase
 C:1-319/Product: ornithine carbamoyltransferase #status experimental <MAT>
 F:12-310/Domains: aspartate/ornithine carbamoyltransferase homology <ACT>

Query Match 64.2%; Score 34; DB 1; Length 319;
 Best Local Similarity 77.8%; Pred. No. 24;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 SFKDMQLG 10
 Db 79 SQKDLQLG 87

RESULT 7
 mosc protein - Rhizobium meliloti (strain L5-30)
 C:Species: Rhizobium meliloti
 C:Date: 08-Sep-1995 #sequence_revision 08-Sep-1995 #text_change 17-Mar-2000
 C:Accession: D53308
 R:Murphy, P.J.; Trenz, S.P.; Grzemska, W.; De Bruijn, F.J.; Schell, J.
 J. Bacteriol. 175, 5193-5204, 1993
 A:Title: The Rhizobium meliloti rhizopine mos locus is a mosaic structure facilitating
 A:Reference number: A53308; MUID:93352426
 A:Accession: D53308
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-407 <MUR>
 A:Cross-references: GB:L17071; NID:q310301; P1DN:AAA26303.1; P1D:q310305
 C:Genetics:
 A:Gene: mosC
 C:Superfamily: cynx protein

Query Match 64.2%; Score 34; DB 2; Length 407;
 Best Local Similarity 66.7%; Pred. No. 32;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 YSFKNMQLG 9
 Db 61 YSFNIMQLG 69

RESULT 8
 F84552
 similar to MLO proteins from H. vulgare [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: F84552
 R:Liu, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
 K.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanden, S.E.; Umayam, L.; Tallon
 euss, D.; Nieman, W.C.; White, O.; Elsen, J.A.; Salzberg, S.L.; Fraser, C.M.; Vent
 Nature 402, 761-766, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: AB4420; MUID:20083487
 A:Accession: F84552
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-574 <STO>
 A:Cross-references: GB:AE002093; NID:q41914369; P1DN:AA032905.1; GSPDB:GM00139
 C:Genetics:
 A:Gene: At2g17480
 A:Map position: 2

Query Match 64.2%; Score 34; DB 2; Length 574;
 Best Local Similarity 70.0%; Pred. No. 46;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 YSFKNMQLG 10
 Db 194 YSFLTMQLG 203

RESULT 9

OKRTA1
acyl-CoA oxidase (EC 1.3.3.6) chain A, peroxisomal splice form I - rat
N:Contains: acyl-CoA oxidase component B: acyl-CoA oxidase component C
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 10-Dec-1999
C:Accession: A29328
R: Miyazawa, S.; Hayashi, H.; Hijioka, M.; Ishii, N.; Furuta, S.; Kagamiyama, H.; Osumi, J. Biol. Chem. 262, 8131-8137, 1987
A:Title: Complete nucleotide sequence of cDNA and predicted amino acid sequence of rat a
A:Reference number: A92649; MUID:87250404
A:Accession: A29328
A:Molecule type: mRNA
A:Residues: 1-661 <MT>
A:Cross-references: GB:J02752; NID:g202677; PIDN:AAA40666.1; PID:g202678
C:Comment: This FAD enzyme, located in peroxisomes, catalyzes the oxygen-specific oxidat
xide.
C:Comment: This enzyme is a mixture of different combinations of three polypeptide chain
e PIR:OKRTA2.
C:Comment: The peroxisomes are eukaryote subcellular organelles that generate and degrad
ic reactions as well as the enzymes involved are biochemically and immunologically diff
C:Superfamily: acyl-CoA oxidase
C:Keywords: alternative splicing; FAD; fatty acid oxidation; flavoprotein; oxidoreductas
F:1-438/Product: acyl-CoA oxidase, peroxisomal splice form I, chain B #status predicted
F:439-661/Product: acyl-CoA oxidase, peroxisomal, chain C #status predicted <CCH>
F:659-661/Region: peroxisome/glyoxysome location signal (S-[RKH]-L) motif

Query Match 64.2%; Score 34; DB 1; Length 661;
Best Local Similarity 66.7%; Pred. No. 54;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 YSFKDMQLG 9
: |||||
Db 610 FDFKDMQLG 618

RESULT 10

OKRTA2
acyl-CoA oxidase (EC 1.3.3.6) chain A, peroxisomal splice form II - rat
N:Contains: acyl-CoA oxidase component B: acyl-CoA oxidase component C
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 10-Dec-1999
C:Accession: B29328; I55261
R: Miyazawa, S.; Hayashi, H.; Hijioka, M.; Ishii, N.; Furuta, S.; Kagamiyama, H.; Osumi, J. Biol. Chem. 262, 8131-8137, 1987
A:Title: Complete nucleotide sequence of cDNA and predicted amino acid sequence of rat a
A:Reference number: A92649; MUID:87250404
A:Accession: B29328
A:Molecule type: mRNA
A:Residues: 1-661 <MT>
R: Osumi, T.; Ishii, N.; Miyazawa, S.; Hashimoto, T.
J. Biol. Chem. 262, 8138-8143, 1987
A:Title: Isolation and structural characterization of the rat acyl-CoA oxidase gene.
A:Reference number: I55261; MUID:87250405
A:Accession: I55261
A:Status: translated from GB/EMBL/DBD3
A:Molecule type: DNA
A:Residues: 1-36 <RES>
A:Cross-references: GB:J02753; NID:g202679; PIDN:AAA40667.1; PID:g554410
C:Comment: This FAD enzyme, located in peroxisomes, catalyzes the oxygen-specific oxidat
xide.
C:Comment: This enzyme is a mixture of different combinations of three polypeptide chain
e PIR:OKRTA1.
C:Comment: The peroxisomes are eukaryote subcellular organelles that generate and degrad
ic reactions as well as the enzymes involved are biochemically and immunologically diff
C:Genetics:
A:Gene: aCOA
C:Superfamily: acyl-CoA oxidase
C:Keywords: alternative splicing; FAD; fatty acid oxidation; flavoprotein; oxidoreductas
F:1-438/Product: acyl-CoA oxidase, peroxisomal splice form II, chain B #status predicted
F:439-661/Product: acyl-CoA oxidase, peroxisomal, chain C #status predicted <CCH>

F:659-661/Region: peroxisome/glyoxysome location signal (S-[RKH]-L) motif

Query Match 64.2%; Score 34; DB 1; Length 661;
Best Local Similarity 66.7%; Pred. No. 54;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 YSFKDMQLG 9
: |||||
Db 610 FDFKDMQLG 618

RESULT 11

H83190
probable outer membrane protein PA3648 [imported] - Pseudomonas aeruginosa (strain PA
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83190
R: Stover, C.K.; Pham, X.Q.; Ervin, A.L.; Mizoguchi, S.D.; Watters, P.; Hickey, M.J.;
Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidig, K.; L
Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337
A:Accession: H83190
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-797 <STO>
A:Cross-references: GB:AE004784; GB:AE004091; NID:g9949799; PIDN:AGC07036.1; GSPDB:GN
C:Genetics:
A:Experimental source: strain PA01
A:Gene: PA3648
C:Superfamily: protective surface antigen D-15

Query Match 64.2%; Score 34; DB 2; Length 797;
Best Local Similarity 75.0%; Pred. No. 66;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 FFKDMQLGR 10
: |||||
Db 71 FDFKDMQLGR 78

RESULT 12

GB4790
probable ABC transporter [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: GB4790
R: Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanhaken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: GB4790
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1413 <STO>
A:Cross-references: GB:AEU02093; NID:g4056482; PIDN:AAC98048.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g37280
A:Map position: 2

Query Match 64.2%; Score 34; DB 2; Length 1413;
Best Local Similarity 62.5%; Pred. No. 1,2e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 3 FFKDMQLGR 10
: |||||
Db 452 FFKDMQLGR 459

RESULT 13

DB1328

probable riboflavin synthase (EC 2.5.1.9) alpha chain Cj1218c [imported] - Campylobacter
C:Species: Campylobacter jejuni

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 08-Sep-2000

C:Accession: DB1328

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelley, J.M.; Churcher, C.; Basham, D.; Chilling

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barr

Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hy

A:Reference number: A81250; MUID:20150912

A:Accession: DB1328

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-203 <PAR>

A:Cross-references: GB:AL139077; GB:AL111168; NID:96968444; PIDN:CAB73472.1; PID:969686

A:Experimental source: serotype O2, strain NCTC 11168

A:Gene: ribA; Cj1218c

C:Superfamily: riboflavin synthase alpha chain

C:Keywords: transferase

Query Match

Best Local Similarity 62.3%; Score 33; DB 2; Length 203;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 FKDMQLGR 10

DB 160 FKDYQVGR 167

RESULT 14

T02418

hypothetical protein At2g23730 [imported] - Arabidopsis thaliana
N:Alternate names: hypothetical protein F27L4.9

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 16-Feb-2001

C:Accession: T02418; C84628

R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,

submitted to the EMBL Data Library, May 1998

A:Description: Arabidopsis thaliana chromosome II BAC F27L4 genomic sequence.

A:Reference number: Z14658

A:Accession: T02418

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-216 <ROU>

A:Cross-references: EMBL:AC004482; NID:93152602; PIDN:AAC17089.1; PID:93152610

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.L.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Clinton, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487

A:Accession: C84628

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-216 <STO>

A:Cross-references: GB:AE002093; NID:93152610; PIDN:AAC17089.1; CSPDB:GN00139

C:Genetics:

A:Gene: At2g23730; F27L4.9

A:Map position: 2

A:Introns: 71/3; 172/3

C:Superfamily: Arabidopsis hypothetical protein F27L4.9

Query Match 62.3%; Score 33; DB 2; Length 216;
Best Local Similarity 70.0%; Pred. NO. 26;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 YSFKDMQLGR 10

DB 114 YSWADMQLVR 123

RESULT 15

G75167

hypothetical protein PAB0339 - Pyrococcus abyssi (strain Orsay)

C:Species: Pyrococcus abyssi

C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000

C:Accession: G75167

R:Anonymous; Genoscope

submitted to the EMBL Data Library, July 1999

A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s

A:Reference number: A75001

A:Accession: G75167

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-225 <KAW>

A:Cross-references: GB:AJ248284; GB:AL096836; NID:95457730; PIDN:CAB49422.1; PID:9545

A:Experimental source: strain Orsay

C:Genetics:

A:Gene: PAB0339

C:Superfamily: conserved hypothetical protein M01163

Query Match

Best Local Similarity 62.3%; Score 33; DB 2; Length 225;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 YSFKDMQL 8

DB 141 YVFKDMEL 148

Search completed: February 27, 2002, 11:45:12
Job time: 691 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:42:38 ; Search time 78.39 seconds
(without alignments)
4.677 Million cell updates/sec

Title: US-09-446-109a-2
Perfect score: 53
Sequence: 1 YSFKDMLCR 10

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	73.6	1676	1 COS_HUMAN	P01031 Homo sapien
2	34	64.2	319	1 OTCA_BACSU	P18186 bacillus su
3	34	64.2	407	1 MOSC_RHME	Q07609 rhizobium m
4	34	64.2	560	1 YML21_ARATH	Q22757 arabidopsis
5	34	64.2	661	1 CAOP_RAT	P07872 rattus norv
6	33	62.3	305	1 ATNB_TORCA	P05029 torpedo cal
7	33	62.3	349	1 RECA_BACHD	Q93647 bacillus ha
8	33	62.3	350	1 XRP2_HUMAN	Q75695 homo sapien
9	33	62.3	385	1 LEF3_NPVAC	P14553 autographa
10	33	62.3	388	1 RECA_STRPN	P30758 streptococc
11	33	62.3	439	1 RECA_ARATH	Q39199 arabidopsis
12	33	62.3	541	1 601M_HAETN	P44973 haemophilus
13	33	62.3	678	1 YIG7_YEAST	P40514 saccharomyc
14	33	62.3	1738	1 YCF1_EPIVI	Q00383 epifagus vi
15	32	60.4	239	1 RNC_HELPJ	Q92182 helicobacte
16	32	60.4	239	1 RNC_HELPJ	P56118 helicobacte
17	32	60.4	266	1 ETRX_STPAU	P01552 staphylococ
18	32	60.4	341	1 RECA_LACLA	Q39486 lactococcus
19	32	60.4	343	1 RECA_CAMJE	P21440 campylobact
20	32	60.4	356	1 MURB_BUCAI	P71153 buchnera ap
21	32	60.4	468	1 YOKI_BACSU	P44571 bacillus su
22	32	60.4	567	1 NU2M_HANMI	P48906 hansenula w
23	32	60.4	584	1 FHAC_BORPE	P35077 bordetella
24	32	60.4	4128	1 FRKD_HUMAN	P78527 Homo sapien
25	31	58.5	74	1 COSA_BOVIN	P12082 bos taurus
26	31	58.5	137	1 COPI_DICDI	P54706 dictyosteli
27	31	58.5	162	1 CBP6_YEAST	P07253 saccharomyc
28	31	58.5	215	1 3F_DICDI	P13475 dictyosteli
29	31	58.5	264	1 WCG3_NPVAC	P16091 autographa
30	31	58.5	266	1 WCG3_NPVAC	Q90173 bombyx mori
31	31	58.5	318	1 ONCA_TAETA	P22080 taenia taen
32	31	58.5	318	1 RECA_BACFR	P22841 bacteroides
33	31	58.5	357	1 YE28_HELPJ	Q9ZJ14 helicobacte

34	31	58.5	357	1 YE28_HELPJ	Q25970 helicobacte
35	31	58.5	382	1 YCB1_MESVI	Q9M033 mesocytigma
36	31	58.5	597	1 MBHL_CITFR	Q46046 citrobacter
37	31	58.5	3635	1 LMA5_MOUSE	Q61001 mus musculu
38	31	58.5	3712	1 ACVS_CERAC	P25464 cephalospor
39	31	58.5	3712	1 LMA_DROME	Q00174 drosophila
40	30	56.6	71	1 Y16K_BPT4	P39243 bacterioph
41	30	56.6	74	1 COSA_PIG	P01032 sus scrofa
42	30	56.6	99	1 YC65_GUTR	Q78422 guillardi
43	30	56.6	111	1 YIFE_ECOLI	P27827 escherichia
44	30	56.6	134	1 H3_ENTHT	Q06196 entamoeba h
45	30	56.6	135	1 H31_SCHPO	P09988 schizosacch

ALIGNMENTS

RESULT	1	ALIGNMENTS
COS_HUMAN	STANDARD:	PRT: 1676 AA.
ID	COS_HUMAN	
AC	P01031:	
IP	21-JUL-1986 (Rel. 01, Created)	
PT	01-DEC-1992 (Rel. 24, Last sequence update)	
DT	20-AUG-2001 (Rel. 40, Last annotation update)	
DE	COMPLETE C5 PRECURSOR [CONTAINS: C5A ANAPHYLATOXIN].	
UN	C5.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
CX	NCBI_TaxID=9606;	
LN	[1]	
PP	SEQUENCE FROM N.A.	
XX	MEDLINE=91079575; PubMed=1984448;	
KA	Haviland D.L., Haviland J.C., Fleischer D.T., Hunt A., Wetsel R.A.;	
KT	"Complete cDNA sequence of human complement pro-C5. Evidence of	
KT	truncated transcripts derived from a single copy gene.";	
RL	J. Immunol. 146:362-368(1991).	
RN	[2]	
RP	SEQUENCE OF 412-1676 FROM N.A.	
XX	MEDLINE=88209511; PubMed=3365401;	
KA	Wetsel R.A., Lemons R.S., Lebeau M.M., Barnum S.R., Noack D.,	
KT	Tack B.F.;	
KT	"Molecular analysis of human complement component C5: localization of	
KT	the structural gene to chromosome 9.";	
RL	Biochemistry 27:1474-1482(1988).	
RN	[3]	
RP	SEQUENCE OF 412-902 FROM N.A.	
XX	MEDLINE=85130937; PubMed=2579066;	
KA	Lundwall A.B., Wetsel R.A., Kristensen T., Whitehead A.S.,	
KA	Woods D.E., Ogden R.C., Colten H.R., Tack B.F.;	
KT	"Isolation and sequence analysis of a cDNA clone encoding the fifth	
KT	complement component.";	
RL	J. Biol. Chem. 260:2108-2112(1985).	
RN	[4]	
RP	SEQUENCE OF 678-751.	
XX	MEDLINE=79005687; PubMed=690134;	
KA	Fernandez H.N., Hugli T.E.;	
KT	"Primary structural analysis of the polypeptide portion of human C5a	
KT	anaphylatoxin. Polypeptide sequence determination and assignment of	
KT	the oligosaccharide attachment site in C5a.";	
RL	J. Biol. Chem. 253:6955-6964(1978).	
LN	[5]	
RE	SEQUENCE OF 678-751 FROM N.A.	
XX	MEDLINE=9114547; PubMed=1996961;	
KA	Bonnasack J.F., Mollison K.W., Buko A.M., Ashworth J.C., Hill H.R.;	
KT	"Group B streptococci inactivate complement component C5a by enzymic	
KT	cleavage at the C-terminus.";	
RL	Biochem. J. 273:635-640(1991).	
RN	[6]	
RP	STRUCTURE BY NMR OF C5A.	
XX	MEDLINE=88309754; PubMed=3408713;	
KA	Zuiderweg E.R.P., Mollison K.W., Henkin J., Carter G.W.;	
KT	"Sequence-specific assignments in the 1H NMR spectrum of the human	

RT Inflammatory protein C5a.";
 RL Biochemistry 27:3568-3580(1988).
 RN (77)
 RP STRUCTURE BY NMR OF C5A.
 RX MEDLINE-89207527; PubMed-2784981;
 RA Zunderweg E.R.P., Nettesheim D.G., Mollison K.W., Carter G.W.;
 RT "tertiary structure of human complement component C5a in solution
 from nuclear magnetic resonance data.";
 RL Biochemistry 28:172-185(1989).
 RN (81)
 RP STRUCTURE BY NMR OF C5A.
 RX MEDLINE-89274164; PubMed-2730871;
 RA Zunderweg E.R.P., Fesik S.W.;
 RT "heteronuclear three-dimensional NMR spectroscopy of the inflammatory
 protein C5a.";
 RL Biochemistry 28:2387-2391(1989).
 RN (91)
 RP STRUCTURE BY NMR OF C5A.
 RX MEDLINE-97160477; PubMed-9007977;
 RA Zhang X., Boyar W., Galakatos N., Gonnella N.C.;
 RT "Solution structure of a unique C5a semi-synthetic antagonist:
 implications in receptor binding.";
 RL Protein Sci. 6:65-72(1997).
 RN (10)
 RP STRUCTURE BY NMR OF C5A.
 RX MEDLINE-97332508; PubMed-9188742;
 RA Zhang X., Boyar W., Toch M.J., Menogole L., Gonnella N.C.;
 RT "Structural definition of the C5a C terminus by two-dimensional
 nuclear magnetic resonance spectroscopy.";
 RL Protein 28:261-267(1997).
 CC -1- FUNCTION: ACTIVATION OF C5 BY A C5 CONVERTASE INITIATES THE
 CC SPONTANEOUS ASSEMBLY OF THE LATE COMPLEMENT COMPONENTS, C5-C9,
 CC INTO THE MEMBRANE ATTACK COMPLEX. C5B HAS A TRANSIENT BINDING SITE
 CC FOR C6. THE C5B-C6 COMPLEX IS THE FOUNDATION UPON WHICH THE LYTIC
 CC COMPLEX IS ASSEMBLED.
 CC -1- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C5,
 CC INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR
 CC PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND
 CC BASOPHILIC LEUKOCYTES. C5A ALSO STIMULATES THE LOCOMOTION OF
 CC POLYMORPHONUCLEAR LEUKOCYTES (CHEMOKINESIS) AND DIRECT THEIR
 CC MIGRATION TOWARD SITES OF INFLAMMATION (CHEMOTAXIS).
 CC -1- SUBUNIT: C5 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 BASIC
 CC RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE
 CC BOND. C5 CONVERTASE ACTIVATES C5 BY CLEAVING THE ALPHA CHAIN,
 CC RELEASING C5A ANAPHYLATOXIN & GENERATING C5B (BETA CHAIN + ALPHA
 CC CHAIN).
 CC -1- SIMILARITY: TO C3, C4 AND ALPHA-2-MACROGLOBULIN.
 CC -1- SIMILARITY: CONTAINS 1 ANAPHYLATOXIN-LIKE DOMAIN.
 CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 855
 CC ONWARD DUE TO THE PRESENCE OF AN ALU REPEAT.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M57729; AAS1925.1; -;
 DR EMBL: M65134; AAS1856.1; -;
 DR PIR: A40075; C5HU.
 DR PIR: S15121; S15121.
 DR PDB: 1KJ5; 15-MAY-97.
 DR PDB: 1CFN; 17-SEP-97.
 DR MIM: 120900; -;
 DR InterPro: IPR002890; A2M_N.
 DR InterPro: IPR001599; Alpha_2_macroglobulin.
 DR InterPro: IPR000020; Anaphylatoxin.
 DR InterPro: IPR001840; Anaphylatoxin.
 DR InterPro: IPR001134; Neutrin_C.
 DR Pfam: PF00207; A2M; 1.

LR Pfam; PF01835; A2M_N; 1.
 DR Pfam; PF01821; ANATO; 1.
 DR Pfam; PF01759; NTR; 1.
 DR PRINTS; PRO0004; ANAPHYLATOXN.
 DR SMART; SM00104; ANATO; 1.
 DR PROSITE; PS00477; ALPHA_2-MACROGLOBULIN; FALSE_NEG.
 DR PROSITE; PS01177; ANAPHYLATOXIN_1; 1.
 DR PROSITE; PS01178; ANAPHYLATOXIN_2; 1.
 KW Complement pathway: Complement alternate pathway: Glycoprotein;
 KW Plasma: Membrane attack complex: Cytolysis; Inflammatory response;
 KW Signal: Polymorphism: 3d-structure.
 FT SIGNAL 1 18
 FT CHAIN 1 673
 FT PROPEP 674 677
 FT CHAIN 678 1676
 FT PEPTIDE 678 751
 FT CHAIN 752 1676
 FT DOMAIN 698 732
 FT DISULFID 698 724
 FT DISULFID 699 731
 FT DISULFID 711 732
 FT CARBOHYD 741 741
 FT CARBOHYD 911 911
 FT CARBOHYD 1115 1115
 FT CARBOHYD 1630 1630
 FT VARIANT 518 518
 FT SEQUENCE 1676 AA; 188331 MW; 87DCA655FE977D19 CRC64;
 SO /FTID=VAR_001996.
 QY 2 SFKMDQLGR 10
 DB 743 SHKMDQLGR 751
 RESULT 2
 ID OTCA_BACSU STANDARD; PRT: 319 AA.
 AC P18186;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ORNITHINE CARBAMOYLTRANSFERASE, ANABOLIC (EC 2.1.3.3) (OTCASE).
 GN ARGF.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 CX NCBI_TaxID=1423;
 GN (1)
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RC STRAIN=168 / EWG50;
 RX MEDLINE-90356402; PubMed-2117745;
 RA Mountain A., Smith M.C.M., Baumberg S.;
 RT "Nucleotide sequence of the Bacillus subtilis argf gene encoding
 RT ornithine carbamoyltransferase.";
 RL Nucleic Acids Res. 18:4594-4594(1990).
 RN (12)
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE-94297722; PubMed-8025667;
 RA O'Reilly M., Devine K.M.;
 RT "Sequence and analysis of the cltC gene biosynthetic operon argC-F
 RT from Bacillus subtilis.";
 RL Microbiology 140:1023-1025(1994).
 RN (13)
 RP SEQUENCE OF 82-319 FROM N.A.
 RC STRAIN=C0741;
 RX MEDLINE-97474245; PubMed-9335269;
 KK Ogura M., Ohshiro Y., Hiroo S., Tanaka T.;
 GN

```

RT      *A new Bacillus subtilis gene, med, encodes a positive regulator of
RT      conk.
RL      J. Bacteriol. 179:6244-6253(1997).
CC      -1- CATALYTIC ACTIVITY: CARBAMOYL PHOSPHATE + ORNITHINE - CITRULLINE
CC      + ORTHOPHOSPHATE.
CC      -1- PATHWAY: ARGININE BIOSYNTHESIS.
CC      -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC      -1- SIMILARITY: BELONGS TO THE ATFCASES/OTFCASES FAMILY.
CC      -1- This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X53360; CAAB3744.1; -
DR      EMBL; 226919; CAAB1542.1; -
DR      EMBL; 299109; CAB12966.1; -
DR      EMBL; 299110; CAB12982.1; -
DR      EMBL; D86376; BAA22924.1; -
DR      PIR; S11000; OMS.
DR      PIR; S38434; S38434.
DR      HSSP; Q51742; IAI5.
DR      SubList; BC10197; argF.
DR      InterPro; IPR002029; Carbonyltransf_asor.
DR      Pfam; PF00185; OTCase; 1.
DR      Pfam; PF02729; OTCase_N; 1.
DR      PRINTS; PR00100; AOTCase.
DR      PRINTS; PR00102; OTCase.
DR      PROSITE; PS00097; CARBAMOYLTRANSFERASE; 1.
KW      Transferase; Arginine biosynthesis; Complete proteome.
SQ      SEQUENCE 319 AA; 34663 MW; E3BE18A5A9BDB9DE CRC64;

Query Match          64.2%; Score 34; DB 1; Length 319;
Best Local Similarity 77.8%; Pred. No. 11;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      2 SFKDMOLCR 10
Db      79 SOKDLQICR 87
      1 11:1111
      1 11:1111

RESULT 3
MOSC_RHIME          STANDARD;      PRT;      407 AA.
AC      007609; Q52891;
DT      01-FEB-1995 (Rel. 31, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      MEMBRANE PROTEIN MOSC.
GN      MOSC.
OS      Rhizobium meliloti (Sinorhizobium meliloti).
OC      Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC      Rhizobiaceae; Sinorhizobium.
OX      NCBI_TaxID=382;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-L5-30;
RX      MEDLINE-93352426; PubMed-8349559;
RA      Murphy P.J., Trenz S.P., Grzesinski W., de Bruijn F.J., Schell J.;
RT      "The Rhizodium meliloti rhizopine mos locus is a mosaic structure
RT      facilitating its symbiotic regulation.";
RL      J. Bacteriol. 175:5193-5204(1993).
RN      [2]
RN      SEQUENCE FROM N.A.
RC      STRAIN-RM220-3;
RX      MEDLINE-96004467; PubMed-7551036;
RA      Rao J.P., Grzesinski W., Murphy P.J.;
RT      "Rhizodium meliloti lacking mosA synthesizes the rhizopine scyllo-
RT      inosamine in place of 3-O-methyl-scyllo-inosamine.";

```

```

RL      Microbiology 141:1683-1690(1995).
CC      -1- FUNCTION: MAY BE A MEMBRANE TRANSPORT PROTEIN THAT COULD EITHER
CC      TRANSPORT A PRECURSOR FOR RHIZOPINE BIOSYNTHESIS INTO BACTERIODS
CC      OR THE FINISHED PRODUCT FROM THE BACTERIODS.
CC      -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC      -1- ALTERNATIVE PRODUCTS: TWO PROTEINS, A 41 AND A 43 KDA ISOFORMS,
CC      MAY BE PRODUCED BY THE USE OF IN-FRAME ALTERNATIVE INITIATION
CC      CODONS.
CC      -1- TISSUE SPECIFICITY: EXPRESSED IN NODULES.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; L17071; AAA26303.1; -
DR      EMBL; U23753; AAA91314.1; -
KW      Transmembrane; Transport; Alternative initiation.
ET      CHAIN 1 407
ET      CHAIN 20 407
ET      INT_MET 20 20
ET      TRANSMEM 31 51
ET      TRANSMEM 69 89
ET      TRANSMEM 109 129
ET      TRANSMEM 157 177
ET      TRANSMEM 186 206
ET      TRANSMEM 225 245
ET      TRANSMEM 255 275
ET      TRANSMEM 290 310
ET      TRANSMEM 316 336
ET      TRANSMEM 347 367
ET      TRANSMEM 377 397
ET      TRANSMEM 64 64
ET      VARIANT 71 71
ET      VARIANT 90 90
ET      VARIANT 100 100
ET      VARIANT 164 164
ET      VARIANT 174 174
ET      VARIANT 184 184
ET      VARIANT 212 214
ET      VARIANT 219 219
ET      VARIANT 234 254
ET      VARIANT 254 254
ET      CONFLICT 369 384
SQ      SEQUENCE 407 AA; 43216 MW; 09C8F4C56AC7624F CRC64;

Query Match          64.2%; Score 34; DB 1; Length 407;
Best Local Similarity 66.7%; Pred. No. 14;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy      1 YSEKDMOLG 9
Db      61 YSEDAEILG 69
      1 1111:11
      1 1111:11

RESULT 4
MOSC_ARATH          STANDARD;      PRT;      560 AA.
AC      022757;
DT      15-JUL-1999 (Rel. 38, Created)
DT      15-JUL-1999 (Rel. 38, Last sequence update)
DT      20-AUG-2001 (Rel. 40, Last annotation update)
DE      HYPOTHETICAL MLO-LIKE PROTEIN F536.23.
GN      F536.23.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

```

```

OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Bueli C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA Moffat K.S., Cronin L.A., Shen M., Vanaken S.E., Umayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Cready T.H.,
RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA Mierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana."
RL Nature 402:761-768(1999).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE MLO FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AC002329; AAB86524.1;
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 12 32 POTENTIAL.
FT TRANSMEM 47 67 POTENTIAL.
FT TRANSMEM 93 113 POTENTIAL.
FT TRANSMEM 182 202 POTENTIAL.
FT TRANSMEM 305 325 POTENTIAL.
FT TRANSMEM 397 417 POTENTIAL.
SQ SEQUENCE 560 AA: 63247 MW: 65496 P15A6AF9 CRC64:
OY 1 YSFKDMQLGR 10
DB 194 YSFLTMLGR 203

```

```

RA Osumi T., Ishii N., Miyazawa S., Hashimoto T.;
RT "Isolation and structural characterization of the rat acyl-CoA
RT oxidase gene."
RL J. Biol. Chem. 262:8138-8143(1987).
CC -1- FUNCTION: CATALYZES THE DESATURATION OF VERY LONG CHAIN ACYL-COAS
CC TO 2-TRANS-ENOL-COAS.
CC -1- CATALYTIC ACTIVITY: ACYL-CoA + O(2) = TRANS-2,3-DEHYDROACYL-CoA +
CC H(2)O(2) (ACTS ON COA DERIVATIVES OF FATTY ACIDS WITH CHAIN LENGTH
CC FROM 8 TO 18).
CC -1- COFACTOR: FAD.
CC -1- PATHWAY: INITIAL STEP OF THE PEROXISOMAL FATTY ACID BETA-OXIDATION
CC SYSTEM.
CC -1- SUBUNIT: THE ENZYME CONTAINS THREE COMPONENTS A,B AND C. THE
CC LATTER TWO BEING PRODUCED FROM THE FIRST BY A PROTEOLYTIC
CC CLEAVAGE.
CC -1- SUBCELLULAR LOCATION: PEROXISOMAL.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1 (SHOWN HERE) AND 2: ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: HIGH, TO OTHER ACYL-COENZYME A OXIDASES; LOW, TO ACYL-
CC COA DEHYDROGENASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J02752; AAA40666.1; -
DR EMBL: J02753; AAA40667.1; -
DR PIR: A29328; OXRTAL.
DR PIR: B29328; OXRTA2.
DR InterPro: IPR002655; ACOX.
DR Pfam: PF01756; ACOX.1.
DR PROSITE: PS00342; MICROBODIES_CTER; 1.
KW Oxidoreductase; Fatty acid metabolism; Flavoprotein; FAD;
KW Alternative splicing.
FT CHAIN 1 661 ACYL-CoA OXIDASE I PEROXISOMAL, A CHAIN.
FT CHAIN 1 438 ACYL-CoA OXIDASE I PEROXISOMAL, B CHAIN.
FT CHAIN 439 661 ACYL-CoA OXIDASE I PEROXISOMAL, C CHAIN.
FT SITE 659 661 MICROBODY TARGETING SIGNAL.
FT VARSPLIC 90 133 KLVLANFVEPVGLVSWHTPTLLQGTAAQDEKMRPSOEL
FT QII -> NSVIRGHPEPIIDILHMFPTLLHQATVAEQDER
FT FMPANLEIT (IN ISOFORM 2).
SQ SEQUENCE 661 AA: 74678 MW: 24816 P10P066C29E CRC64:
OY 1 YSFKDMQLG 9
DB 610 FDFKDMTLG 618

```

```

RESULT 5
CAOP_RAT STANDARD: PRT: 661 AA.
ID CAOP_RAT P11354:
AC P07872: P11354:
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ACYL-COENZYME A OXIDASE I, PEROXISOMAL (EC 1.3.3.6) (PALMITOYL-CoA
DE OXIDASE) (AOX).
GN ACOX1 OR ACOX.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=87250404; PubMed=3036800;
RA Miyazawa S., Hayashi H., Hijikata M., Ishii N., Furuta S.,
RA Katamiyama H., Osumi T., Hashimoto T.;
RT "Complete nucleotide sequence of cDNA and predicted amino acid
RT sequence of rat acyl-CoA oxidase."
RL J. Biol. Chem. 262:8131-8137(1987).
RN 12
RP SEQUENCE OF 1-36 FROM N.A.
RX MEDLINE=87250405; PubMed=3036801;

```

```

RESULT 6
ATNB_TORCA STANDARD: PRT: 305 AA.
ID ATNB_TORCA P05029:
AC P05029:
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SODIUM/POTASSIUM-TRANSPORTING ATPASE BETA CHAIN (SODIUM/POTASSIUM-
DE DEPENDENT ATPASE BETA SUBUNIT).
OS Torpedo californica (Pacific electric ray).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Squalae; Hypnosqualae; Pristiogaster; Batoidae;
OC Torpediniformes; Torpedinidae; Torpedinidae; Torpedo.
OX NCBI_TaxID=7787;
RN 11
RP SEQUENCE FROM N.A.

```


RA MEDLINE=86136540; PubMed=3005037;
 RA Noguchi S., Noda M., Takahashi H., Kawakami K., Ohta T., Nagano K.,
 RA Hirose T., Iiyama S., Kawamura M., Numa S.;
 RT "Primary structure of the beta-subunit of Torpedo californica (Na⁺ +
 RT K⁺)-ATPase deduced from the cDNA sequence.";
 RL FEBS Lett. 196;315-320(1986).
 RN (1)
 RP DISULFIDE BOND IN 215-278.
 RX MEDLINE=89322300; PubMed=2546555;
 RA Kellaris K.V.;
 RT "Identification of a disulfide between cysteine 214 and cysteine 277
 RT in the beta subunit of native (Na⁺ + K⁺)-ATPase.";
 RL Biochem.-Biophys. Res. Commun. 162:64-70(1989).
 CC -1- FUNCTION: THIS IS THE NON-CATALYTIC COMPONENT OF THE ACTIVE
 CC ENZYME, WHICH CATALYZES THE HYDROLYSIS OF ATP COUPLED WITH THE
 CC EXCHANGE OF NA AND K IONS ACROSS THE PLASMA MEMBRANE. THE BETA
 CC SUBUNIT REGULATES, THROUGH ASSEMBLY OF ALPHA/BETA HETERODIMERS,
 CC THE NUMBER OF SODIUM PUMPS TRANSPORTED TO THE PLASMA MEMBRANE.
 CC -1- SUBUNIT: COMPOSED OF THREE SUBUNITS: ALPHA (CATALYTIC), BETA
 CC AND GAMMA.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE NA⁺/K⁺ AND H⁺ ATPASES BETA CHAIN
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X03471; CAA27188.1; -
 DR PIR: A33625; PMRYNB.
 DR PIR: A33291; A33291.
 DR InterPro: IPR000402; Na_K_beta.
 DR Pfam: PF00287; Na_K-ATPase; 1.
 DR PROSITE: PS00390; ATPASE_NA_K_BETA_1; 1.
 DR PROSITE: PS00391; ATPASE_NA_K_BETA_2; 1.
 KW Sodium/potassium transport; Transmembrane; Glycoprotein;
 KW Signal-anchor.
 FT DOMAIN 1 32 CYTOPLASMIC (POTENTIAL).
 FT TRANSM 33 53 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 54 305 EXTRACELLULAR (POTENTIAL).
 FT DISULFID 127 150 BY SIMILARITY.
 FT DISULFID 160 176 BY SIMILARITY.
 FT DISULFID 215 278 BY SIMILARITY.
 FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 159 159 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
 FT CARBOHYD 194 194 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
 FT CARBOHYD 267 267 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
 SQ SEQUENCE 305 AA; 34672 MW; BCI3191F8D597563 CnC64;
 Query Match 62.3%; Score 33; DB 1; Length 305;
 Best Local Similarity 70.0%; Pred. No. 16;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 YSFKMOQLCR 10
 DB 287 YSEKDRSLCR 296
 RESULT 7
 RECA_BACHD STANDARD: PRT: 349 AA.
 ID RECA_BACHD
 AC O9KAA7;
 DT 20-AUG-2001 (Rel. 40; Created)
 DT 20-AUG-2001 (Rel. 40; Last sequence update)
 DT 20-AUG-2001 (Rel. 40; Last annotation update)
 DE RECA PROTEIN (RECOMBINASE A).
 GN RECA OR BH2383.

OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=86665;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 RT halodurans and genomic sequence comparison with Bacillus subtilis.";
 RL Nucleic Acids Res. 28:4317-4331(2000).
 CC -1- FUNCTION: CAN CATALYZE THE HYDROLYSIS OF ATP IN THE PRESENCE OF
 CC SINGLE-STRANDED DNA. THE ATP-DEPENDENT UPTAKE OF SINGLE-STRANDED
 CC DNA BY DUPLEX DNA, AND THE ATP-DEPENDENT HYBRIDIZATION OF
 CC HOMOLOGOUS SINGLE-STRANDED DNAs. IT INTERACTS WITH LEXA CAUSING
 CC ITS ACTIVATION AND LEADING TO ITS AUTOCATALYTIC CLEAVAGE (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE RECA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AP001515; BAB06102.1; -
 DR InterPro: IPR003593; AAA.
 DR InterPro: IPR001553; RECA.
 DR Pfam: PF00154; RECA; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00321; RECA_1; 1.
 DR PROSITE: PS00162; RECA_2; 1.
 DR PROSITE: PS00163; RECA_3; 1.
 KW DNA damage; DNA recombination; SOS response; ATP-binding; DNA-binding;
 KW Complete proteome.
 FT NP-BIND 64 ATP (BY SIMILARITY).
 FT SEQUENCE 349 AA; 37914 MW; EFB6B57675F8104 CnC64;
 Query Match 62.3%; Score 33; DB 1; Length 349;
 Best Local Similarity 60.0%; Pred. No. 19;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 1 YSFKMOQLCR 10
 DB 289 YSFNDRQLQ 298
 RESULT 8
 XRP2_HUMAN
 ID XRP2_HUMAN STANDARD: PRT: 350 AA.
 AC O75695;
 DT 30-MAY-2000 (Rel. 39; Created)
 DT 30-MAY-2000 (Rel. 39; Last sequence update)
 DT 20-AUG-2001 (Rel. 40; Last annotation update)
 DE XRP2 PROTEIN.
 GN XRP2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A., AND VARIANTS XLRP SER-6 DEL AND HIS-118.
 RC TISSUE=Brain;
 RX MEDLINE=98361157; PubMed=9697592;
 RX Schwahn U., Lenzner S., Dong J., Feil S., Hinzmann B.,
 RX van Duijnhoven G., Kirschner R., Hemberger M., Bergen A.A.B.,

RA Rosenberg T., Pinckers A.J.L.G., Fundele R., Rosenthal A.,
 RA Cremers F.P.M., Ropers H.-H., Berger W.:
 RT "Positional cloning of the gene for x-linked retinitis pigmentosa 2.";
 RN Nat. Genet. 19:327-332(1998).
 [2]
 RP VARIANT XLRP HIS-118.
 RX MEDLINE-99192368; PubMed-10090907;
 RA Hardcastle A.J., Thibault D.L., Van Maldergem L., Saha B.K., Jay M.,
 RA Plant C., Taylor R., Bird A.C., Bhattacharya S.:
 RT "Mutations in the RP2 gene cause disease in 10% of families with
 RT familial x-linked retinitis pigmentosa assessed in this study.";
 RL Am. J. Hum. Genet. 64:1210-1215(1999).
 [3]
 RP VARIANTS: XLRP SER-6 DEL AND HIS-118.
 RX MEDLINE-99451197; PubMed-10520237;
 RA Rosenberg T., Schwahn U., Feil S., Berger W.:
 RT "Genotype-phenotype correlation in x-linked retinitis pigmentosa 2
 RT (RP2).";
 RL Ophthalmic Genet. 20:161-172(1999).
 [4]
 RP VARIANT XLRP ARG-253.
 RX MEDLINE-20098121; PubMed-10634633;
 RA Mada Y., Nakazawa M., Abe T., Tamai M.:
 RT "A new Leu253Arg mutation in the RP2 gene in a Japanese family with
 RT x-linked retinitis pigmentosa.";
 RL Invest. Ophthalmol. Vis. Sci. 41:290-293(2000).
 CC -1- DISEASE: DEFECTS IN RP2 ARE RESPONSIBLE FOR X-LINKED RETINITIS
 CC PIGMENTOSA-2 (XLRP-2 OR RP2). CHARACTERIZED BY CONSTRUCTION OF THE
 CC VISUAL FIELDS, NIGHT BLINDNESS, AND FUNDUS CHANGES, INCLUDING
 CC 'BONE CORPUSCLE' LUMPS OF PIGMENT.
 CC -1- SIMILARITY: BELONGS TO THE TRGC FAMILY.
 CC -1- DATABASE: NAME=Mutations of the RP2 gene;
 CC NOTE=Retina International's Scientific Newsletter;
 CC WWW="http://www.retina-international.com/sci-news/rp2mut.htm".
 [5]
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AJ007590; CA07577.1; -.
 DR MIM: 312600; -.
 KW Vision; Retinitis pigmentosa; Disease mutation.
 FT VARIANT 6 6 MISSING (IN RP2).
 FT VARIANT 108 108 /FTid=VAR_008497.
 FT VARIANT 118 118 C->G (IN RP2).
 FT VARIANT 118 118 /FTid=VAR_008498.
 FT VARIANT 118 118 R->H (IN RP2).
 FT VARIANT 253 253 /FTid=VAR_008499.
 FT VARIANT 253 253 L->R (IN RP2).
 FT SEQUENCE 350 AA: 39642 MW: 687FCFB8C53086C5 CRC64:
 SQ
 Query Match 62.3%; Score 33; DB 1; Length 350;
 Best Local Similarity 55.6%; Pred. NO. 19;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DE LATE EXPRESSION FACTOR 3.
 GN LEF-3.
 OS Autographa californica nuclear polyhedrosis virus (AcMNPV).
 CC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
 CC Nucleopolyhedrovirus.
 OX NCBI_TaxID=46015;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP STRAIN=LI;
 RC MEDLINE-93353600; PubMed-8350397;
 RX LI Y., Passarelli A.L., Miller L.K.:
 PA "Identification, sequence, and transcriptional mapping of lef-3, a
 RT baculovirus gene involved in late and very late gene expression.";
 RL J. Virol. 67:5260-5268(1993).
 [2]
 RP SEQUENCE FROM N.A.
 RP STRAIN=C6;
 RC MEDLINE-94303173; PubMed-8030224;
 RA Ayres M.D., Howard S.C., Kuzio J., Lopez-Ferber M., Possee R.D.:
 RT "The complete DNA sequence of Autographa californica nuclear
 RT polyhedrosis virus.";
 RL Virology 202:586-605(1994).
 CC -1- FUNCTION: REQUIRED FOR LATE AND VERY LATE GENE EXPRESSION. LEF-3
 CC COULD BE A SINGLE STRANDED DNA-BINDING PROTEIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L18873; AAA02964.1; -.
 DR EMBL: L2858; AAA66697.1; -.
 DR PIR: A40677; A40677.
 KW Early protein; Transcription regulation; DNA-binding.
 SQ SEQUENCE 385 AA: 44551 MW: 9A25ECD7BA7FBDE1 CRC64:
 Query Match 62.3%; Score 33; DB 1; Length 385;
 Best Local Similarity 66.7%; Pred. NO. 21;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Cy 2 SFKDMOLC 10
 Eb 73 SFKDMOLC 81

RESULT 10
 ID RECA_STRPN STANDARD: PRT: 388 AA.
 AC P30758;
 UF 01-JUL-1993 (Rel. 26, Last sequence update)
 UF 01-JUL-1993 (Rel. 26, Last sequence update)
 UF 20-AUG-2001 (Rel. 40, Last annotation update)
 DE RECA PROTEIN (RECOMBINASE A).
 GN RECA OR SP1940.
 OS Streptococcus pneumoniae.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 CC Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP STRAIN=R800;
 RC MEDLINE-93117122; PubMed-1475203;
 RA Martin B., Ruellan J.M., Angulo J., Devoret R., Claverys J.-P.:
 RT "Identification of the recA gene of Streptococcus pneumoniae.";
 RL Nucleic Acids Res. 20:6412-6412(1992).
 [2]
 RP SEQUENCE FROM N.A.
 RP STRAIN=TIGR4;
 RC MEDLINE-11463916;
 KX

```

RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Dinkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiloul S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.:
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RT Science. 293:498-506(2001).
RN [3]
RP SEQUENCE OF 1-11 FROM N.A.
RC STRAIN-R6X:
RX MEDLINE=95095991; PubMed=7798154;
RA Pearce B.J., Naughton A.M., Campbell E.A., Masure H.R.:
RT "The rec locus, a competence-induced operon in Streptococcus
RT pneumoniae."
RL J. Bacteriol. 177:86-93(1995).
CC -1- FUNCTION: CAN CATALYZE THE HYDROLYSIS OF ATP IN THE PRESENCE OF
CC SINGLE-STRANDED DNA, THE ATP-DEPENDENT UPTAKE OF SINGLE-STRANDED
CC DNA BY DUPLEX DNA, AND THE ATP-DEPENDENT HYBRIDIZATION OF
CC HOMOLOGOUS SINGLE-STRANDED DNAs. IT INTERACTS WITH LEXA CAUSING
CC ITS ACTIVATION AND LEADING TO ITS AUTOCATALYTIC CLEAVAGE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- INDUCTION: BY COMPETENCE AND DAMAGE.
CC -1- SIMILARITY: BELONGS TO THE RECA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z17307; CA87895.1; -
DR EMBL: Z34303; CA84072.1; -
DR EMBL: AE007484; AAK76008.1; -
DR EMBL: L36131; AAC09383.1; -
DR PIR: S30219; S30219.
DR HSSP: P03017; 2REB.
DR TIGR: SP1940; -
DR InterPro: IPR001553; RecA.
DR Pfam: PF00154; recA.1.
DR PRINTS: PR00142; RECA.
DR PROSITE: PS00321; RECA_1; 1.
DR PROSITE: PS50162; RECA_2; 1.
DR PROSITE: PS50163; RECA_3; 1.
KW DNA damage; DNA recombination; SOS response; ATP-binding; DNA-binding.
FT NP_BIND 79 86 ATP (BY SIMILARITY).
SQ SEQUENCE 368 AA; 41949 MW; 40F5DA5F15E1BC83 CRC64;
QY 1 YSFKDMQGR 10
Db 308 YSKDEKIQG 317
Query Match 62.3%; Score 33; DB 1; Length 388;
Best Local Similarity 50.0%; Pred. No. 21;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

```

```

CC Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
CC Eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RX MEDLINE=94302159; PubMed=8029344;
RA Blinet M.N., Osman M., Jagendorf A.T.:
RT "Genomic nucleotide sequence of a gene from Arabidopsis thaliana
RT encoding a protein homolog of Escherichia coli RecA."
RL Plant Physiol. 103:673-674(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA Theologis A., Vysotskaya V.S., Osborne B.I., Schwartz J.R.,
RA Federspiel N.A., Kwan A., Tortum M., Yu G., Oji O., Araujo R.,
RA Chung E., Dewar K., Dietrich F., Ecker J.R., Marzall A., Oefner P.,
RA Davis R.W.:
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 2-439 FROM N.A.
RX MEDLINE=92390390; PubMed=1518831;
RA Cerutti H.D., Osman M., Grandoni P., Jagendorf A.T.:
RT "A homolog of Escherichia coli RecA protein in plastids of higher
RT plants."
PL Proc. Natl. Acad. Sci. U.S.A. 89:8068-8072(1992).
CC -1- FUNCTION: INVOLVED IN RECOMBINATION ABILITY AND DNA STRAND
CC TRANSFER ACTIVITY.
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST STROMA.
CC -1- SIMILARITY: BELONGS TO THE RECA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L15229; AAA61781.1; -
DR EMBL: AC002986; AAC17051.1; -
DR EMBL: M98039; AAA32855.1; -
DR HSSP: P03017; 2REB.
DR InterPro: IPR003593; AAA.
DR InterPro: IPR001553; RecA.
DR Pfam: PF00154; recA.1.
DR PRINTS: PR00142; RECA.
DR SMART: SM00382; AAA.1.
DR PROSITE: PS00321; RECA_1; 1.
DR PROSITE: PS50162; RECA_2; 1.
DR PROSITE: PS50163; RECA_3; 1.
KW DNA damage; DNA recombination; ATP-binding; DNA-binding; Chloroplast;
KW Transil peptide.
FT TRANSIT 1 52 CHLOROPLAST (POTENTIAL).
FT CHAIN 53 439 DNA REPAIR PROTEIN RECA.
FT NP_BIND 142 149 ATP (POTENTIAL).
SQ SEQUENCE 439 AA; 47733 MW; 3545E98F7B37E15 CRC64;
QY 1 YSFKDMQGR 10
Db 369 YSYEDQRLQG 378
Query Match 62.3%; Score 33; DB 1; Length 439;
Best Local Similarity 50.0%; Pred. No. 24;
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

```

```

RESULT 11
RECA_ARATH STANDARD: PRT: 439 AA.
AC Q39199: Q39200:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA REPAIR PROTEIN RECA, CHLOROPLAST PRECURSOR.
CN RECA OR YUP8H12R.33.
OS Arabidopsis thaliana (Mouse-ear cress).

```

```

RESULT 12
60IM_HAEIN STANDARD: PRT: 541 AA.
ID 60IM_HAEIN
AC P44973;

```

```

DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE 60 KDA INNER-MEMBRANE PROTEIN HOMOLOG.
GN H1101.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischman R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA McLachlan A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,
RA Fine L.D., Fitchman J.L., Fuhrmann J.L., Geophagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT Influenzae Rd."
RL Science 269:496-512(1995).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE OXAL / 60 KDA IMP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32781; AAC22663.1; -
DR TIGR: H11001; -
DR InterPro: IPR001708; 60KD_inner_MP.
DR Pfam: PF02096; 60KD_IMP.1.
DR PRINTS: PRO0701; 60KDINNERMP.
KW Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 6 26 POTENTIAL.
FT TRANSMEM 313 333 POTENTIAL.
FT TRANSMEM 337 357 POTENTIAL.
FT TRANSMEM 417 437 POTENTIAL.
FT TRANSMEM 495 515 POTENTIAL.
SQ SEQUENCE 541 AA; 61239 MW; 9406398B334F4EDE CRC64;

Query Match 62.3%; Score 33; DB 1; Length 541;
Best Local Similarity 85.7%; Pred. No. 30;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

LN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / AB972;
RA Bartell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagers K., Jones M.,
RA Louis E., Lye G., Moulé S., Moulé T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skellon J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z38060; CA86156.1; -
DR PIR: S48412; S48412.
DR SCD: S0001329; Y1L067C.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 119 139 POTENTIAL.
FT TRANSMEM 245 265 POTENTIAL.
FT TRANSMEM 317 337 POTENTIAL.
FT TRANSMEM 340 360 POTENTIAL.
FT TRANSMEM 371 391 POTENTIAL.
FT TRANSMEM 405 425 POTENTIAL.
FT TRANSMEM 443 463 POTENTIAL.
FT TRANSMEM 475 495 POTENTIAL.
FT TRANSMEM 519 539 POTENTIAL.
SQ SEQUENCE 678 AA; 75482 MW; B237BB728D029F0C CRC64;

Query Match 62.3%; Score 33; DB 1; Length 678;
Best Local Similarity 66.7%; Pred. No. 38;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

RT *Function and evolution of a minimal plastid genome from a
RT nonphotosynthetic parasitic plant.*
RL Proc. Natl. Acad. Sci. U.S.A. 89:10648-10652(1992).
CC -1- FUNCTION: NOT YET KNOWN.
CC -1- SIMILARITY: BELONGS TO THE YCF1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X61368; CAA43644.1; -
DR EMBL: M81884; AAA65870.1; -
DR PIR: S16720; S16720.
DR PIR: S20614; S20614.
DR Mendel: 5139; EPIV1:ycf1.1.
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 1738 AA; 208318 MW; 65C63F63BDC8364B CRC64;

Query Match 62.3%; Score 33; DB 1; Length 1738;
Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSPKDMOL 8
| | | | |
Db 997 YSPKEMDL 1004

RESULT 15
RNC_HELPJ
ID RNC_HELPJ STANDARD: PRT: 239 AA.
AC Q9ZLH2:
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RIBONUCLEASE III (EC 3.1.26.3) (RNASE III).
CN RNC OR JHP0607.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Molr D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummlin P.J., Garuso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT *Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.*;
RL Nature 397:176-180(1999).
CC -1- FUNCTION: DIGESTS DOUBLE-STRANDED RNA. INVOLVED IN THE PROCESSING
CC OF RIBOSOMAL RNA PRECURSORS AND OF SOME MRNAs (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ENDONUCLEOLYTIC CLEAVAGE TO 5'-PHOSPHO-
CC MONOESTER.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: CONTAINS 1 DBM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE RNASE III FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001492; AAD06188.1; -

```

```

DR InterPro: IPR001159; DS_RBD.
DR InterPro: IPR000999; RNase_3.
DR Pfam: PF00035; dsrm: 1.
DR Pfam: PF00636; Ribonuclease_3; 1.
DR SMART: SM00358; DSRM: 1.
DR SMART: SM00355; RIBOC: 1.
DR PROSITE: PS50137; DS_RBD: 1.
DR PROSITE: PS00517; RNASE_3_1; 1.
DR PROSITE: PS50142; RNASE_3_2; 1.
KW Hydrolase; Nuclease; Endonuclease; RNA-binding; Complete proteome.
FT DOMAIN 219 DBM.
SQ SEQUENCE 239 AA; 27275 MW; 197DF81F5BE701D5 CRC64;

Query Match 60.4%; Score 32; DB 1; Length 239;
Best Local Similarity 75.0%; Pred. No. 20;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSPKDMOL 8
| | | | |
Db 27 YSPKDKRL 34

```

Search completed: February 27, 2002, 11:42:39
Job time: 538 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: February 27, 2002, 11:49:59 : Search time 281.76 Seconds
(Without alignments)
5.191 Million cell updates/sec

Title: US-09-446-109a-2
Perfect score: 53
Sequence: 1 YSFKDMQLGR 10

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 473505 seqs, 14627329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database:

SPREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protist:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	73.6	571	2	Q9PN32 campylobact
2	36	67.9	397	11	Q64194 rattus sp.
3	36	67.9	560	10	Q49141 arabidopsis
4	36	67.9	560	10	Q50057 arabidopsis
5	36	67.9	611	2	Q9X552 streptomyce
6	36	67.9	754	10	Q49144 arabidopsis
7	36	67.9	791	10	Q50073 arabidopsis
8	36	67.9	791	10	Q49139 arabidopsis
9	36	67.9	791	10	Q49137 arabidopsis
10	35	66.0	175	2	Q92671 borrelia tu
11	35	66.0	283	5	Q93252 caenorhabdi
12	35	66.0	331	13	Q90324 cyprinus ca
13	35	66.0	385	5	Q9VGH6 diosiphila
14	35	66.0	833	5	Q20374 caenorhabdi
15	34	64.2	354	2	Q44384 agrobacteri
16	34	64.2	450	2	Q99MA4 staphylococ
17	34	64.2	574	10	Q9SHM0 arabidopsis
18	34	64.2	797	2	Q9HX14 pseudomonas
19	34	64.2	864	2	Q68395 thauera aro

20	34	64.2	1413	10	Q9ZUT8 arabidopsis
21	33	62.3	188	2	P95667 clostridium
22	33	62.3	203	2	Q9PN77 campylobact
23	33	62.3	216	10	Q64829 arabidopsis
24	33	62.3	225	1	Q9VIC7 pyrococcus
25	33	62.3	226	1	Q9HH77 pyrococcus
26	33	62.3	236	1	Q59330 pyrococcus
27	33	62.3	242	12	Q84406 paramecium
28	33	62.3	265	4	Q9H514 homo sapien
29	33	62.3	314	2	Q9CN39 pasteurella
30	33	62.3	377	10	Q9L784 arabidopsis
31	33	62.3	385	12	Q92432 bombix mori
32	33	62.3	533	10	Q9FT12 arabidopsis
33	33	62.3	569	10	Q9FKY5 arabidopsis
34	33	62.3	773	5	Q9V365 diosiphila
35	33	62.3	2093	5	Q9XZY8 leishmania
36	32.5	61.3	592	4	Q95697 homo sapien
37	32.5	61.3	1294	4	Q9UQP3 homo sapien
38	32	60.4	82	2	Q99246 streptococc
39	32	60.4	115	2	Q9PIS9 campylobact
40	32	60.4	136	5	Q9N2N8 campylobact
41	32	60.4	170	2	P94482 bacillus su
42	32	60.4	217	2	Q67705 aquilex aeo
43	32	60.4	241	2	Q9K5T9 bacillus ha
44	32	60.4	320	10	Q9LZ22 arabidopsis
45	32	60.4	402	5	Q9TXV1 caenorhabdi

ALIGNMENTS

RESULT 1
Q9PN32 PRELIMINARY: PRT: 571 AA.
ID Q9PN32: K/C Q9PN32: 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NI/FE-HYDROGENASE LARGE SUBUNIT.
CN HYDB OR CUI266.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
CC Campylobacter.
CX NCBI_Taxid:197;
RN [1]
RP SEQUENCE FROM N.A.
KC STRAIN-NCCTC 11168;
RX MERLINE-20150912: PubMed-10688204;
RA Parkhill J., Wren B.W., Mungall K., Kelsey J.M., Churcher C., Basham D., Chillingworth T., Davies R.M., Feltham T., Holtroyd S., Jagers K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W., Quail M.A., Rajandream M.A., Rutherford K.M., Van Vliet A.H.M., Whitehead S., Barrett B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervariable sequences..";
KL Nature 403:665-668(2000).
DR EMBL: AL139077: CAB73520.1: -;
DR InterPro: IPR001501: Nifese_Hases.
DR Pfam: PF00374: Nifese_Hases.1;
DR PROSITE: PS00507: NI_HGENSE_L_1: 1;
DR PROSITE: PS00508: NI_HGENSE_L_2: 1;
KW Complete proteome.
SQ SEQUENCE 571 AA: 63570 MW: 47886046B876C23F CRC64;

Query Match 73.6%; Score 39; DB 2; Length 571;
Best Local Similarity 60.0%; Pred. No. 7.6;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Caps 0;

QY 1 YSFKDMQLGR 10
DB 282 YTFKPEQICR 291

```

RESULT 2
ID 064194 PRELIMINARY: PRT: 397 AA.
AC 064194:
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE LYSOSOMAL ACID LIPASE.
OS Rattus sp.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10118.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96129534; PubMed=8576647;
RA Nakagawa H., Matsubara S., Kuriyama M., Yoshidome H., Fujiyama J.,
RA Yoshida H., Osame M.;
RT "Cloning of rat lysosomal acid lipase cDNA and identification of the
RT mutation in the rat model of Wolman's disease.";
RL J. Lipid Res. 36:2212-2218(1995).
CC -1- SIMILARITY: TO ALPHA/BETA HYDROLASE FOLD.
DE EMBL: S81497; AAB36043.2;
DE MODIFICATION METHYLASE (EC 2.1.1.73) (CYTOSINE-SPECIFIC
DE METHYLTRANSFERASE).
CN Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702.
RN [1]
RP SEQUENCE FROM N.A.
RX TRANSPOSON-EVELKNLEVEL RETROTRANSPOSON:
RA Henikoff S., Comal L.;
RL Genetics 0:0-0(1998).
CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + DNA CYTOSINE = S-
CC ADENSYL-L-HOMOCYSTEINE + DNA 5-METHYLCYTOSINE.
DE -1- SIMILARITY: TO C-5 CYTOSINE-SPECIFIC DNA METHYLASE FAMILY.
DE EMBL: AF039372; AAC02667.1;
DR Mendel: 24694; Arach:2780;24694.
DR InterPro: IPR001025; BAH.
DR InterPro: IPR001525; C5_DNA_meth.
DR InterPro: IPR000953; Chromo.
DR Pfam: PF01426; BAH; 1.
DR Pfam: PF00385; Chromo; 1.
DR Pfam: PF00145; DNA_methylase; 1.
DR PRINTS: PR00105; CSMETFRASE.
DR PRINTS: PR00504; CHROMODOMAIN.
DR SMART: SM00439; BAH; 1.
DR SMART: SM00298; CHROMO_1; 1.
DR PROSITE: PS00598; CHROMO_1; 1.

```

```

DR PROSITE: PS00503; CHROMO_2; 1.
KW Methyltransferase; Transferase.
SQ SEQUENCE 560 AA; 63514 MW; 16AFD7FF53CC354B CRC64;

Query Match
Best Local Similarity 67.9%; Score 36; DB 10; Length 560;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 FKDMQLGR 10
DB 537 FKDLQVGR 544

RESULT 4
ID 050057 PRELIMINARY: PRT: 560 AA.
AC 050057:
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE MODIFICATION METHYLASE (EC 2.1.1.73) (CYTOSINE-SPECIFIC
DE METHYLTRANSFERASE).
CN Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702.
RN [1]
RP SEQUENCE FROM N.A.
RX TRANSPOSON-EVELKNLEVEL RETROTRANSPOSON:
RA Henikoff S., Comal L.;
RL Genetics 0:0-0(1998).
CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + DNA CYTOSINE = S-
CC ADENSYL-L-HOMOCYSTEINE + DNA 5-METHYLCYTOSINE.
DE -1- SIMILARITY: TO C-5 CYTOSINE-SPECIFIC DNA METHYLASE FAMILY.
DE EMBL: AF039373; AAC02668.1;
DR Mendel: 24765; Arach:2780;24765.
DR InterPro: IPR001025; BAH.
DR InterPro: IPR001525; C5_DNA_meth.
DR InterPro: IPR000953; Chromo.
DR Pfam: PF01426; BAH; 1.
DR Pfam: PF00385; Chromo; 1.
DR Pfam: PF00145; DNA_methylase; 1.
DR PRINTS: PR00504; CHROMODOMAIN.
DR SMART: SM00439; BAH; 1.
DR SMART: SM00298; CHROMO_1; 1.
DR PROSITE: PS00598; CHROMO_1; 1.
DR PROSITE: PS0013; CHROMO_2; 1.
KW Methyltransferase; Transferase.
SQ SEQUENCE 560 AA; 63484 MW; E5757F9E49669BFO CRC64;

Query Match
Best Local Similarity 75.0%; Score 36; DB 10; Length 560;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 FKDMQLGR 10
DB 537 FKDLQVGR 544

RESULT 5
ID 09X5S2 PRELIMINARY: PRT: 611 AA.
AC 09X5S2:
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE MUCD.
KW MUCD.

```


OS Streptomyces lavendulae.
 OC Bacteria: Firmicutes: Actinobacteria: Actinobacteridae;
 OC Actinomycetales: Streptomyicinae: Streptomycetaceae: Streptomyces.
 OX NCBI_TaxID=1914;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRRL 2564;
 RX MEDLINE=99201491; PubMed=10099135;
 RA Mo Y.Q., Varoglu M., Sherman D.H.;
 RT "Molecular characterization and analysis of the biosynthetic gene
 RT cluster for the antitumor antibiotic mitomycin C from Streptomyces
 RT lavendulae NRRL 2564."
 RL Chem. Biol. 6:251-263(1999).
 DR EMBL: AF127374; AAD32727.1; -;
 SO SEQUENCE 611 AA; 68499 MW; 350B189E5D19EC64 CRC64;

Query Match 67.9% Score 36; DB 2; Length 611;
 Best Local Similarity 60.0% Pred. No. 35;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 YSKDMOLGR 10
 DB 442 YAFKDPYGR 451

RESULT 6
 ID 049144 PRELIMINARY: PRT: 754 AA.
 AC 049144;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE MODIFICATION METHYLASE (EC 2.1.1.73) (CYTOSINE-SPECIFIC
 DE METHYLTRANSFERASE) (FRAGMENT).
 GN CMT1.
 OS Arabidopsis suecica.
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
 OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae;
 OC eustosids II: Brassicales: Brassicaceae: Arabidopsis.
 OX NCBI_TaxID=45249;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ECOTYPE SUE-2(9510);
 RA Henikoff S., Comai L.;
 RL Genetics 0:0-0(1998).
 CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + DNA CYTOSINE = S-
 CC -1- ADENOSYL-L-HOMOCYSTEINE + DNA 5-METHYLCYTOSINE.
 CC -1- SIMILARITY: TO C-5 CYTOSINE-SPECIFIC DNA METHYLASE FAMILY.
 DR EMBL: AF039374; AAC02670.1; -;
 DR Mendel: 24697; Arabid.2780:24697.
 DR InterPro: IPR001025; BAH.
 DR InterPro: IPR001525; C5_DNA_meth.
 DR InterPro: IPR000953; Chromo.
 DR Pfam: PF01426; BAH; 1.
 DR Pfam: PF00385; chromo; 1.
 DR Pfam: PF00145; DNA_methylase; 2.
 DR PRINTS: PRO0105; CSMETFRASE.
 DR PRINTS: PRO0504; CHROMODOMAIN.
 DR SMART: SM00439; BAH; 1.
 DR SMART: SM00298; CHROMO_1.
 DR PROSITE: PS00598; CHROMO_1; 1.
 DR PROSITE: PS50013; CHROMO_2; 1.
 DR MethyLtransferase; Transferase.
 FT NON_TER 754
 SO SEQUENCE 754 AA; 85129 MW; F2322B8B7ACAC011 CRC64;

Query Match 67.9% Score 36; DB 10; Length 754;
 Best Local Similarity 75.0% Pred. No. 45;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 OY 3 FKDMOLGR 10

DB 537 FKDLQVGR 544

RESULT 7
 ID 050073 PRELIMINARY: PRT: 791 AA.
 AC 050073;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE MODIFICATION METHYLASE (EC 2.1.1.73) (CYTOSINE-SPECIFIC
 DE METHYLTRANSFERASE).
 GN CMT1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
 OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae;
 OC eustosids II: Brassicales: Brassicaceae: Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ECOTYPE KL-0;
 RA Henikoff S., Comai L.;
 RL Genetics 0:0-0(1998).
 CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + DNA CYTOSINE = S-
 CC -1- ADENOSYL-L-HOMOCYSTEINE + DNA 5-METHYLCYTOSINE.
 CC -1- SIMILARITY: TO C-5 CYTOSINE-SPECIFIC DNA METHYLASE FAMILY.
 DR EMBL: AF039366; AAC02659.1; -;
 DR Mendel: 24771; Arabid.2780:24771.
 DR InterPro: IPR001025; BAH.
 DR InterPro: IPR001525; C5_DNA_meth.
 DR InterPro: IPR000953; Chromo.
 DR Pfam: PF01426; BAH; 1.
 DR Pfam: PF00385; chromo; 1.
 DR Pfam: PF00145; DNA_methylase; 2.
 DR PRINTS: PRO0504; CHROMODOMAIN.
 DR SMART: SM00439; BAH; 1.
 DR SMART: SM00298; CHROMO_1.
 DR PROSITE: PS00598; CHROMO_1; 1.
 DR PROSITE: PS50013; CHROMO_2; 1.
 DR MethyLtransferase; Transferase.
 FT NON_TER 791
 SO SEQUENCE 791 AA; 89186 MW; C21043E553409EDF CRC64;

Query Match 67.9% Score 36; DB 10; Length 791;
 Best Local Similarity 75.0% Pred. No. 47;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 3 FKDMOLGR 10
 DB 537 FKDLQVGR 544

RESULT 8
 ID 049139 PRELIMINARY: PRT: 791 AA.
 AC 049139;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE MODIFICATION METHYLASE (EC 2.1.1.73) (CYTOSINE-SPECIFIC
 DE METHYLTRANSFERASE).
 GN CMT1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
 OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae;
 OC eustosids II: Brassicales: Brassicaceae: Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ECOTYPE ND-1 AND ECOTYPE ND-0;
 RA Henikoff S., Comai L.;

```

RL  Genetics 0:0-0(1998).
CC  -i- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + DNA CYTOSINE - S-
CC  ADEHYDROXYL-L-HOMOCYSTEINE + DNA 5-METHYLCYTOSINE.
CC  -i- SIMILARITY: TO C-5 CYTOSINE-SPECIFIC DNA METHYLASE FAMILY.
DR  EMBL: AF039370; AAC02663.1; -.
DR  EMBL: AF039369; AAC02662.1; -.
DR  Mendel: 24692; Arach:2780;24692.
DR  InterPro: IPR001025; BAH.
DR  InterPro: IPR001525; C5_DNA_meth.
DR  InterPro: IPR000953; Chromo.
DR  Pfam: PF01426; BAH; 1.
DR  Pfam: PF00385; Chromo; 1.
DR  Pfam: PF00145; DNA_methylase; 2.
DR  PRINTS: PR00504; CHROMODOMAIN.
DR  SMART: SM00439; BAH; 1.
DR  SMART: SM00298; CHROMO; 1.
DR  PROSITE: PS00598; CHROMO_1; 1.
DR  PROSITE: PS50013; CHROMO_2; 1.
KW  Methyltransferase; Transferase.
SQ  SEQUENCE 791 AA: 89174 MW: E39CFACE5249562B CRC64:

- QY 3 FKMOLGR 10
      111:111
Db 537 FKDLQVGR 544

Query Match          67.9%; Score 36; DB 10; Length 791;
Best Local Similarity 75.0%; Pred. No. 47;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

RESULT 9
ID 049137 PRELIMINARY; PRT: 791 AA.
AC 049137;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE MODIFICATION METHYLASE (EC 2.1.1.73) (CYTOSINE-SPECIFIC
DE METHYLTRANSFERASE).
GN CMT1 OR F23A5.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ECOTYPE COL-0;
RA Henikoff S., Comai L.;
RL Genetics 0:0-0(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Schwartz J.R., Yu C., Tortum M., Lenz C., Liu S., Lee J.M., Li J.,
RA Gonzalez A., Liu A., Liu K., Sakano H., Koo T., Pham P., Vaysberg M.,
RA Howang B., Chin C., Choi E., Chio J., Altati H., Brooks S., Chao Q.,
RA Conn L., Conway A., Hansen N., Johnson-Hopson C., Khan S., Kim C.,
RA Lam B., Nguyen M., Palm C., Shinn P., Tambunga G., Davis R.W.,
RA Ecker J.R., Federspiel N.A., Theologis A.;
RT "Arabidopsis thaliana chromosome 1 BAC F23A5 sequence.";
RT submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Theologis A.;
RL submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Theologis A.;
RL submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC -i- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + DNA CYTOSINE - S-

```

```

CC  ADEHYDROXYL-L-HOMOCYSTEINE + DNA 5-METHYLCYTOSINE.
CC  -i- SIMILARITY: TO C-5 CYTOSINE-SPECIFIC DNA METHYLASE FAMILY.
DR  EMBL: AF039367; AAC02660.1; -.
DR  EMBL: AC011713; AAF14662.1; -.
DR  Mendel: 24691; Arach:2780;24691.
DR  InterPro: IPR001025; BAH.
DR  InterPro: IPR001525; C5_DNA_meth.
DR  InterPro: IPR000953; Chromo.
DR  Pfam: PF01426; BAH; 1.
DR  Pfam: PF00385; Chromo; 1.
DR  Pfam: PF00145; DNA_methylase; 2.
DR  PRINTS: PR00105; C5METHYLASE.
DR  PRINTS: PR00504; CHROMODOMAIN.
DR  SMART: SM00439; BAH; 1.
DR  SMART: SM00298; CHROMO; 1.
DR  PROSITE: PS00598; CHROMO_1; 1.
DR  PROSITE: PS50013; CHROMO_2; 1.
KW  Methyltransferase; Transferase.
SQ  SEQUENCE 791 AA: 89218 MW: A5ECFDBC274B215C CRC64:

- QY 3 FKMOLGR 10
      111:111
Db 537 FKDLQVGR 544

Query Match          67.9%; Score 36; DB 10; Length 791;
Best Local Similarity 75.0%; Pred. No. 47;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

RESULT 10
ID 092611 PRELIMINARY; PRT: 175 AA.
AC 092611;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, last annotation update)
DE HYPOTHETICAL 20.5 KDA PROTEIN.
OS Borrelia turicatae.
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=142;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=021;
RA Pennington P.M., Cadavid D., Bunikis K., Norris S., Barbour A.G.;
RA "Interplasmidic arm duplications determine virulence phenotype of the
RA bacterium Borrelia turicatae.";
RT submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=021;
RA Pennington P.M., Cadavid D., Bunikis J., Norris S., Barbour A.G.;
RT "Extensive interplasmidic duplications change the virulence phenotype
RT of the relapsing fever agent Borrelia turicatae.";
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=021;
RA Mo. Microbiol. 43:1120-1132(1999).
RR EMBL: AF049852; AAD15784.1; -.
RR EMBL: AF129434; AAC34130.1; -.
RR Hypothetical protein.
SQ SEQUENCE 175 AA: 20499 MW: AB03B5BC63F17E9 CRC64:

- QY 1 YSPKMOCL 8
      11111:1
Db 3 YSPKMOCL 10

Query Match          66.0%; Score 35; DB 2; Length 175;
Best Local Similarity 75.0%; Pred. No. 14;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

RESULT 11
ID 093252 PRELIMINARY; PRT: 283 AA.

```

```

AC 0931252;
AC 01-FEB-1997 (TREMblrel. 02, Created)
DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE C23H4.4 PROTEIN.
GN C23H4.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pezodermidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Wilkinson J.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kersey J., Kirsten J., Laister N., Latreille P.,
RA Lightbulb J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproatt J., Wohlman P.,
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.
RL Nature 368:32-38(1994).
CC -1 SIMILARITY: BELONGS TO THE CARBOXYLESTERASE TYPE-B FAMILY.
DR EMBL: Z78416; CAB01678.1;
DR HSSP: P37967; 10E3
DR InterPro: IPR002018; Carboxylesterase_B.
DR InterPro: IPR000379; Est_1ip_Thioest_actsite.
DR InterPro: IPR000734; Lipase.
DR Pfam: PF00135; Coesterase; 1.
DR PROSITE: PS00122; CARBOXYLESTERASE_B.1; 1.
DR PROSITE: PS00120; LIPASE_SER; UNKNOWN_1.
KM Hydrolyase.
KW SEQUENCE 283 AA; 31801 MW; 81475B51381AB815 CRC64;
SO

```

```

DR InterPro: IPR000688; Peptidase_C1.
DR InterPro: IPR000169; Thiolprot_act_site.
DR Pfam: PF00112; Peptidase_C1; 1.
DR PRINTS: PR00705; PAPAIN.
DR PROSITE: PS00640; THIOL_PROTEASE_ASN; 1.
DR PROSITE: PS00139; THIOL_PROTEASE_CYS; 1.
DR PROSITE: PS00639; THIOL_PROTEASE_HIS; 1.
KM Hydrolyase; Thiol protease.
KW SEQUENCE 331 AA; 36228 MW; 076DF34D32C1CF82 CRC64;
SO

```

Query Match 66.0%; Score 35; DB 13; Length 331;
Best Local Similarity 70.0%; Pred. No. 29;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

CY 1 YSFKDMQLGR 10
DB 38 YSKDEELGR 47

```

RESULT 13

```

ID 09VQH6 PRELIMINARY; PRT; 385 AA.
AC 09VQH6;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE CG2975 PROTEIN.
GN CG2975.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
CC Ephydroidea; Drosophilidae; Drosophila.
UX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BERKELEY;
KM MEDLINE=20196006; PubMed=10731132;
FX NC
KA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
KA Amaratilake P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.,
KA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
KA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
KA Brandon G.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
KA Man K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
KA Abril J.F., Agbayani A., An H.-J., Andrews-Piankoff C., Baldwin D.,
KA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
KA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
KA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
KA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
KA Cherry J.M., Cavley S., Dahlke C., Davenport I.B., Davies P.,
KA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
KA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
KA Durbin K.J., Evangelista C.C., Ferrara C., Ferreira S., Fleischmann W.,
KA Foster C., Gabriellian A.E., Cary N.S., Gelbart W.M., Glasser K.,
KA Clodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
KA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
KA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
KA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kelchum K.A.,
KA Kimmel B.E., Koulira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
KA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
KA Liu X., Mattei R., McIntosh T.C., McLeod M.P., McPherson A.,
KA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
KA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
KA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
KA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
KA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
KA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
KA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
KA Svirskaas R., Tector R., Turner R., Venter E., Wang A.H., Wang X.,
KA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
KA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
KA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
KA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

```

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*."
 CC Science 287:2185-2195(2000).
 DR EMBL: AE003581; AAF51197.1; -
 DR FlyBase: FBgn0031468; CG2975.
 DR InterPro: IPR002659; Galactosyl.T.
 DR Pfam: PF01762; Galactosyl.T.
 SO SEQUENCE 385 AA; 44593 MW; 62EBDC576CBDFAD2 CRC64;

Query Match 66.0%; Score 35; DB 5; Length 385;
 Best Local Similarity 50.0%; Pred. No. 35;
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 YSFKDMQLG 10
 1 : 1 : 1 : 1 : 1 :
 DB 236 YGEDVELGR 245

RESULT 14

O20374 PRELIMINARY; PRT; 833 AA.
 AC O20374:
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
 DE F43G6.9 PROTEIN.
 GN F43G6.9.
 OS *Caenorhabditis elegans*.
 OC Eukaryota; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 CC Rhabditidae; Peloderiinae; *Caenorhabditis*.
 CX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Swinburne J.;
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL: 283108; CAB05512.1; -
 DR EMBL: 250070; CAB05512.1; JOINED.
 DR EMBL: 250070; CAA90402.1; -
 DR EMBL: 283108; CAA90402.1; JOINED.
 SO SEQUENCE 833 AA; 93671 MW; 5047A926A94E2EE5 CRC64;

Query Match 66.0%; Score 35; DB 5; Length 833;
 Best Local Similarity 75.0%; Pred. No. 81;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 SFKDMQLG 9
 : 1 : 1 : 1 : 1 :
 DB 205 AFKDLQLG 212

RESULT 15

O44384 PRELIMINARY; PRT; 354 AA.
 AC O44384:
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE PERIPLASMIC PROTEIN MOTB PRECURSOR.
 GN MOTB.
 OS *Agrobacterium tumefaciens*.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 CC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 CC Rhizobiaceae; Rhizobium.
 CX NCBI_TaxID=358;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN-15955;
 RX MEDLINE-98143415; PubMed-9484884;
 RA Oger P., Kim K.S., Sackett R.L., Piper K.R., Farrand S.K.;
 RT "Octopine-type T1 plasmids code for a mannopine-inducible dominant-negative allele of trar, the quorum-sensing activator that regulates

RT T1 plasmid conjugal transfer."
 RL Mol. Microbiol. 27:277-288(1998).
 CC -I- FUNCTION: INVOLVED IN MANNOPINE TRANSPORT.
 CC -I- SUBCELLULAR LOCATION: PERIPLASMIC (POTENTIAL).
 DR EMBL: AF242881; AAC18651.1; -
 DR Transport; Periplasmic; Signal; Plasmid.
 KW SIGNAL
 FT SIGNAL 1 29 POTENTIAL.
 FT CHAIN 30 354 INNER MEMBRANE PROTEIN MOTB.
 SO SEQUENCE 354 AA; 38405 MW; 118FC14A422E6A2C CRC64;

Query Match 64.2%; Score 34; DB 2; Length 354;
 Best Local Similarity 66.7%; Pred. No. 51;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 YSFKDMQLG 9
 1 : 1 : 1 : 1 : 1 :
 DB 251 YTFNDQLG 259

Search completed: February 27, 2002, 11:50:01
 Job time: 980 sec

Wed Feb 27 12:14:00 2002

us-09-446-109a-2.ra1

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: February 27, 2002, 11:36:02 ; Search time 132.19 Seconds
(without alignments)
1.702 Million cell updates/sec

Title: US-09-446-109a-2
Perfect score: 53
Sequence: 1 YSFKDWQLGR 10

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, AA.*
1: /cgn2-6/ptodata/2/1aa/5A-COMB.pep.*
2: /cgn2-6/ptodata/2/1aa/5B-COMB.pep.*
3: /cgn2-6/ptodata/2/1aa/6A-COMB.pep.*
4: /cgn2-6/ptodata/2/1aa/6B-COMB.pep.*
5: /cgn2-6/ptodata/2/1aa/PCTUS-COMB.pep.*
6: /cgn2-6/ptodata/2/1aa/Backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	100.0	10	1	US-08-299-285-2
2	53	100.0	10	2	US-08-985-126-2
3	53	100.0	10	5	PCT-US95-11126-2
4	51	96.2	10	1	US-08-299-285-9
5	51	96.2	10	2	US-08-985-126-9
6	51	96.2	10	5	PCT-US95-11126-9
7	47	88.7	10	1	US-08-299-285-5
8	47	88.7	10	1	US-08-299-285-8
9	47	88.7	10	1	US-08-299-285-26
10	47	88.7	10	1	US-08-299-285-27
11	47	88.7	10	1	US-08-299-285-28
12	47	88.7	10	1	US-08-299-285-29
13	47	88.7	10	1	US-08-299-285-30
14	47	88.7	10	1	US-08-299-285-31
15	47	88.7	10	1	US-08-299-285-32
16	47	88.7	10	1	US-08-299-285-33
17	47	88.7	10	1	US-08-299-285-34
18	47	88.7	10	1	US-08-299-285-35
19	47	88.7	10	1	US-08-299-285-36
20	47	88.7	10	1	US-08-299-285-37
21	47	88.7	10	2	US-08-985-126-5
22	47	88.7	10	2	US-08-985-126-8
23	47	88.7	10	2	US-08-985-126-26
24	47	88.7	10	2	US-08-985-126-27
25	47	88.7	10	2	US-08-985-126-28
26	47	88.7	10	2	US-08-985-126-29
27	47	88.7	10	2	US-08-985-126-30

28	47	88.7	10	2	US-08-985-126-31	Sequence 31, Appl
29	47	88.7	10	2	US-08-985-126-32	Sequence 32, Appl
30	47	88.7	10	2	US-08-985-126-33	Sequence 33, Appl
31	47	88.7	10	2	US-08-985-126-34	Sequence 34, Appl
32	47	88.7	10	2	US-08-985-126-35	Sequence 35, Appl
33	47	88.7	10	2	US-08-985-126-36	Sequence 36, Appl
34	47	88.7	10	2	US-08-985-126-37	Sequence 37, Appl
35	47	88.7	10	5	PCT-US95-11126-5	Sequence 5, Appl
36	47	88.7	10	5	PCT-US95-11126-8	Sequence 8, Appl
37	47	88.7	10	5	PCT-US95-11126-26	Sequence 26, Appl
38	47	88.7	10	5	PCT-US95-11126-27	Sequence 27, Appl
39	47	88.7	10	5	PCT-US95-11126-28	Sequence 28, Appl
40	47	88.7	10	5	PCT-US95-11126-29	Sequence 29, Appl
41	47	88.7	10	5	PCT-US95-11126-30	Sequence 30, Appl
42	47	88.7	10	5	PCT-US95-11126-31	Sequence 31, Appl
43	47	88.7	10	5	PCT-US95-11126-32	Sequence 32, Appl
44	47	88.7	10	5	PCT-US95-11126-33	Sequence 33, Appl
45	47	88.7	10	5	PCT-US95-11126-37	Sequence 37, Appl

ALIGNMENTS

RESULT 1
US-08-299-285-2
Sequence 2, Application US/08299285
Patent No. 5696230
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
US-08-299-285-2

Query Match 100.0%; Score 53; DB 1; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSFKDMQLGR 10
|||||
Db 1 YSFKDMQLGR 10

RESULT 2

US-08-985-126-2
Sequence 2, Application US/08985126
Patent No. 5942599
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/985.126
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/299.285
FILING DATE: 31-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: No. 5942599 Relevant
TOPOLOGY: No. 5942599 Relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
US-08-985-126-2

Query Match 100.0%; Score 53; DB 2; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSFKDMQLGR 10
|||||
Db 1 YSFKDMQLGR 10

RESULT 3
PCT-US95-11126-2
Sequence 2, Application PC/TUS9511126
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.

APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman, P.C.
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/11126
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/299,285
FILING DATE: 31-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:
NAME/KEY: Modified-site
LOCATION: 9..10
OTHER INFORMATION: /product="D-Alanine; N-methylated
OTHER INFORMATION: Arginine"
PCT-US95-11126-2

Query Match 100.0%; Score 53; DB 5; Length 10;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSFKDMQLGR 10
|||||
Db 1 YSFKDMQLGR 10

RESULT 4
US-08-299-285-9
Sequence 9, Application US/08299285
Patent No. 5696230
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720

CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
AMTI-SENSE: NO
FRAGMENT TYPE: C-terminal
US-08-299-285-9

Query Match 96.2% Score 51; DB 1; Length 10;
Best Local Similarity 90.0%; Pred. No. 3.9e-05;
Matches 9: Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSFKDMQLGR 10
|||||:|
DB 1 YSFKDMQLGR 10

RESULT 5
US-08-985-126-9
Sequence 9, Application US/08985126
Patent No. 5942599
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/985,126
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994

ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: No. 5942599 Relevant
TOPOLOGY: No. 5942599 Relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
AMTI-SENSE: NO
FRAGMENT TYPE: C-terminal
US-08-985-126-9

Query Match 96.2% Score 51; DB 2; Length 10;
Best Local Similarity 90.0%; Pred. No. 3.9e-05;
Matches 9: Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSFKDMQLGR 10
|||||:|
DB 1 YSFKDMQLGR 10

RESULT 6
PCT-US95-11126-9
Sequence 9, Application PC/TUS9511126
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman, P.C.
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/11126
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/299,285
FILING DATE: 31-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide

HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
PCT-US95-11126-9

Query Match 96.2% Score 51: DB 5: Length 10:
Best Local Similarity 90.0% Pred. No. 3.9e-05:
Matches 9: Conservative 1: Mismatches 0: Indels 0: Gaps 0:

OY 1 YSFKDMOLGR 10
|11111111|
DB 1 YSFKDMOLGR 10

RESULT 7
US-08-299-285-5
Sequence 5, Application US/08299285
Patent No. 5696230
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
MOLECULE TYPE: peptide
TOPOLOGY: not relevant
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
US-08-299-285-5

Query Match 88.7% Score 47: DB 1: Length 10:
Best Local Similarity 90.0% Pred. No. 0.00029:
Matches 9: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

OY 1 YSFKDMOLGR 10
|11111111|
DB 1 YSFKDMOLGR 10

RESULT 8

US-08-299-285-8
Sequence 8, Application US/08299285
Patent No. 5696230
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
MOLECULE TYPE: peptide
TOPOLOGY: not relevant
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:
NAME/KEY: Modified-site
LOCATION: 9
OTHER INFORMATION: "D-Alanine at position 9"
US-08-299-285-8

Query Match 88.7% Score 47: DB 1: Length 10:
Best Local Similarity 90.0% Pred. No. 0.00029:
Matches 9: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

CY 1 YSFKDMOLGR 10
|11111111|
DB 1 YSFKDMOLGR 10

RESULT 9
US-08-299-285-26
Sequence 26, Application US/08299285
Patent No. 5696230
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:

ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:
NAME/KEY: Modified-site
LOCATION: 9..10
OTHER INFORMATION: "D-Alanine at position 9;
OTHER INFORMATION: N-methylated Arginine at position 10"
US-08-299-285-26

Query Match 88.7% Score 47; DB 1; Length 10;
Best Local Similarity 90.0%; Pred. No. 0.00029;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSFKDMQLCR 10
DB 1 YSFKDMQLAR 10

RESULT 10
US-08-299-285-27
Sequence 27, Application US/08299285
Patent No. 5696230
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:
NAME/KEY: Modified-site
LOCATION: 8..9
OTHER INFORMATION: "-Methylated Leucine at
OTHER INFORMATION: position 8; D-Alanine at position 9"
US-08-299-285-27

Query Match 88.7% Score 47; DB 1; Length 10;
Best Local Similarity 90.0%; Pred. No. 0.00029;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSFKDMQLCR 10
DB 1 YSFKDMQLAR 10

RESULT 11
US-08-299-285-28
Sequence 28, Application US/08299285
Patent No. 5696230
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044

INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:
NAME/KEY: Modified-site
LOCATION: 9..10
OTHER INFORMATION: "D-Alanine at position 9;
OTHER INFORMATION: N-methylated Arginine at position 10"
US-08-299-285-28

Query Match 88.7%; Score 47; DB 1; Length 10;
Best Local Similarity 90.0%; Pred. No. 0.00029;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

CY 1 YSFKDMQLGR 10
DB 1 YSFKDMQLAR 10

RESULT 12

US-08-299-285-29
Sequence 29, Application US/08299285
Patent No. 5696230

GENERAL INFORMATION:

APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994
CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044

INFORMATION FOR SEQ ID NO: 29:

SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:

NAME/KEY: Modified-site
LOCATION: 8
OTHER INFORMATION: "N-Methylated, D-Alanine at
OTHER INFORMATION: position 9"
US-08-299-285-29

Query Match 88.7%; Score 47; DB 1; Length 10;
Best Local Similarity 90.0%; Pred. No. 0.00029;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

CY 1 YSFKDMQLGR 10
DB 1 YSFKDMQLAR 10

RESULT 13

US-08-299-285-30
Sequence 30, Application US/08299285
Patent No. 5696230

GENERAL INFORMATION:

APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994
CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:

NAME/KEY: Modified-site

LOCATION: 8..9

OTHER INFORMATION: "N-Methylated leucine at
OTHER INFORMATION: position 8; D-Alanine at position 9"

US-08-299-285-30

Query Match 88.7%; Score 47; DB 1; Length 10;
Best Local Similarity 90.0%; Pred. No. 0.00029;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 YSFKDMOLCR 10
Db 1 YSFKDMOLAR 10

RESULT 14

US-08-299-285-31
Sequence 31, Application US/08299285
Patent No. 5696230
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
TITLE OF INVENTION: C-Terminal Analogs of Csa Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:
NAME/KEY: Modified-site
LOCATION: 7..9
OTHER INFORMATION: "N-methylated Glutamine at
OTHER INFORMATION: position 7; D-Alanine at position 9"
US-08-299-285-31

Query Match 88.7% Score 47: DB 1: Length 10:
Best Local Similarity 90.0% Pred. No. 0.00029;
Matches 9: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

OY 1 YSFKDMOLCR 10
Db 1 YSFKDMOLAR 10

RESULT 15
US-08-299-285-32
Sequence 32, Application US/08299285
Patent No. 5696230
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.

APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
TITLE OF INVENTION: C-Terminal Analogs of Csa Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:
NAME/KEY: Modified-site
LOCATION: 9..10
OTHER INFORMATION: "N-methylated D-Alanine at
OTHER INFORMATION: position 9; N-methylated Arginine at position 10"
US-08-299-285-32

Query Match 88.7% Score 47: DB 1: Length 10:
Best Local Similarity 90.0% Pred. No. 0.00029;
Matches 9: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

OY 1 YSFKDMOLCR 10
Db 1 YSFKDMOLAR 10

Search completed: February 27, 2002, 11:36:02
Job time: 141 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:41:16 : Search time 303.5 Seconds
(without alignments)
1.464 Million cell updates/sec

Title: US-09-446-109a-10
Perfect score: 38
Sequence: 1 FKPLMR 6

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: A_Geneseq_1101.*
1: /SID58/gcgdata/geneseq/AA1980.DAT.*
2: /SID58/gcgdata/geneseq/AA1981.DAT.*
3: /SID58/gcgdata/geneseq/AA1982.DAT.*
4: /SID58/gcgdata/geneseq/AA1983.DAT.*
5: /SID58/gcgdata/geneseq/AA1984.DAT.*
6: /SID58/gcgdata/geneseq/AA1985.DAT.*
7: /SID58/gcgdata/geneseq/AA1986.DAT.*
8: /SID58/gcgdata/geneseq/AA1987.DAT.*
9: /SID58/gcgdata/geneseq/AA1988.DAT.*
10: /SID58/gcgdata/geneseq/AA1989.DAT.*
11: /SID58/gcgdata/geneseq/AA1990.DAT.*
12: /SID58/gcgdata/geneseq/AA1991.DAT.*
13: /SID58/gcgdata/geneseq/AA1992.DAT.*
14: /SID58/gcgdata/geneseq/AA1993.DAT.*
15: /SID58/gcgdata/geneseq/AA1994.DAT.*
16: /SID58/gcgdata/geneseq/AA1995.DAT.*
17: /SID58/gcgdata/geneseq/AA1996.DAT.*
18: /SID58/gcgdata/geneseq/AA1997.DAT.*
19: /SID58/gcgdata/geneseq/AA1998.DAT.*
20: /SID58/gcgdata/geneseq/AA1999.DAT.*
21: /SID58/gcgdata/geneseq/AA2000.DAT.*
22: /SID58/gcgdata/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	92.1	462	22	AAC98875
2	35	92.1	2204	21	AAV51233
3	33	86.8	6	13	AA830577
4	33	86.8	128	21	AA89737
5	33	86.8	128	21	AA89737
6	33	86.8	128	21	AA89737
7	33	86.8	152	19	AAW74783
8	33	86.8	174	21	AA810415
9	33	86.8	174	21	AA852515
10	33	86.8	182	21	AA831129
11	33	86.8	257	21	AA815288

12	33	86.8	274	21	AA815287	Arabidopsis thalia
13	33	86.8	279	21	AA815286	Arabidopsis thalia
14	33	86.8	434	21	AA801208	Corn putative lect
15	32	84.2	629	20	AA835425	Protein involved i
16	32	84.2	920	21	AA870589	Salmonella Pathoge
17	31	81.6	18	16	AA864890	Human C5a receptor
18	31	81.6	19	14	AA837902	Chemotactant C5
19	31	81.6	65	21	AA856373	Arabidopsis thalia
20	31	81.6	65	21	AA860383	Arabidopsis thalia
21	31	81.6	79	22	AA841295	Human polypeptide
22	31	81.6	304	17	AA848714	G-protein coupled
23	31	81.6	304	17	AA802686	G-protein coupled
24	31	81.6	357	22	AA824361	Human EST encoded
25	31	81.6	482	17	AA802151	Human G-protein co
26	31	81.6	482	18	AA823367	Novel G-protein co
27	31	81.6	482	18	AA806540	C5a-like seven tra
28	31	81.6	482	19	AA842376	Homo sapiens C3a r
29	31	81.6	485	21	AA858319	Lung cancer associ
30	31	81.6	488	20	AA836300	Human secreted pro
31	31	81.6	949	22	AA895425	Human protein sequ
32	30	78.9	51	21	AA858536	Arabidopsis thalia
33	30	78.9	90	21	AA853989	Human colon cancer
34	30	78.9	105	21	AA835358	Zea mays protein f
35	30	78.9	129	21	AA840355	Human ORF ORF19
36	30	78.9	160	21	AA814369	Arabidopsis thalia
37	30	78.9	160	21	AA853398	Arabidopsis thalia
38	30	78.9	184	21	AA853397	Arabidopsis thalia
39	30	78.9	185	21	AA814368	Arabidopsis thalia
40	30	78.9	222	20	AA838694	Neisseria meningit
41	30	78.9	222	20	AA838695	Neisseria meningit
42	30	78.9	222	20	AA838696	Neisseria meningit
43	30	78.9	222	20	AA838697	Neisseria gonorrho
44	30	78.9	249	22	AA840207	Human polypeptide
45	30	78.9	260	21	AA843231	Human ORF ORF2995

ALIGNMENTS

RESULT 1

ID AAC98875 standard: Protein: 462 AA.

AC AAC98875:

DT 26-SEP-2001 (first entry)

DE E. coli growth and proliferation related protein sequence SEQ ID NO:345.

EW Escherichia coli: growth; proliferation: microbial; antimicrobial:

KW Bacterial infection; microorganism.

OS Escherichia coli.

PN W0200134810-A2.

PD 17-MAY-2001.

PE 09-NOV-2000: 2000WO-US0950.

PR 09-NOV-1999: 99US-0164415.

RA (ELITRA) ELITRA PHARM INC.

PI Forsyth RA, Ohlsen K, Zyskind J:

NR WPI: 2001-335933/35.

DR N-PSDB: AA84546.

PT Novel nucleic acids that inhibit Escherichia coli proliferation, useful
for screening for homologous genes and for designing expression vectors

PS Claim 19; Page 421-422; 522pp; English.

XX AAH84373 to AAH84499 represent *Escherichia coli* growth and proliferation
 CC related DNA sequences (1). AAH84500 to AAH84670 encode the *E. coli*
 CC growth and proliferation related proteins given in AAG98078 and AAG98830
 CC to AAG98999. (1) can be used as potential targets for the generation of
 CC new antimicrobial agents, and for identification of compounds which
 CC interact with the gene products of (1). In addition the expression of
 CC (1) and the purification of the proteins, the purified proteins can be
 CC used to generate reagents and screen small molecule libraries or other
 CC candidate compound libraries for compounds that can be further developed
 CC to yield novel antimicrobial compounds. In addition, nucleic acid probes
 CC complementary to (1) that are specific for particular species of
 CC microorganisms can be used to identify particular microorganism species
 CC in clinical specimens, therefore, providing a rapid and dependable
 CC method by which to identify the causative agents of a bacterial
 CC infection. Also, antibodies generated against proteins translated from
 CC mRNA transcribed from proliferation-required sequences can also be used
 CC to screen for specific microorganisms that produce such proteins in a
 CC species-specific manner. AAH84371 and AAH84670 represent sequencing
 CC primers used in the isolation of *E. coli* growth and proliferation
 CC related sequence, which are used in an example from the present
 CC invention.

XX Sequence 462 AA:

SO

QY 1 FKPLMR 6
 :|||||
 Db 398 ykplmr 403

Query Match 92.1%; Score 35; DB 22; Length 462;
 Best Local Similarity 83.3%; Pred. No. 48;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

RESULT 2
 AAY51233
 ID AAY51233 standard; Protein: 2204 AA.
 XX
 AC AAY51233:
 XX
 DT 07-APR-2000 (first entry)
 XX
 DE Newcastle disease virus LaSota genome encoded protein 6.
 XX
 KW Avian-paramyxovirus: infection; lentogenic; F protein; vaccine;
 KW respiratory disease; gastrointestinal disease; poultry pathogen;
 KW local immunity.
 XX
 OS Newcastle disease virus.
 XX
 PN WO9966045-A1.
 XX
 PD 23-DEC-1999.
 XX
 PF 17-JUN-1999; 99MO-NL00377.
 XX
 PR 19-JUN-1998; 98EP-0202054.
 XX
 PI (DIEN-) STICHTING DIENST LANDBOUWKUNDIG ONDERZOE.
 XX
 PI Peeters BPH, De Leeuw OS, Koch G, Gielkens ALJ;
 XX
 DR WPI: 2000-106102/09.
 XX
 XX New avian paramyxovirus cDNA, useful for production of vaccine against
 PT Newcastle disease virus -
 XX
 PS Disclosure: Fig 3; 115pp; English.
 XX
 CC This invention describes a novel avian-paramyxovirus cDNA (1) which
 CC comprises a nucleic acid sequence corresponding to the 5' terminal

CC end of the genome of avian-paramyxovirus allowing the generation of
 CC an infectious copy of avian-paramyxovirus. The cell line is useful for
 CC the production of infectious lentogenic NDV (Newcastle Disease virus)
 CC without the addition of exogenous proteolytic activity. Also it is
 CC possible to generate a stable transfected cell line that expresses the
 CC wild-type F protein in the form of a vaccine, especially against
 CC respiratory and/or gastrointestinal diseases. NDV can be easily cultured
 CC to very high titers in embryonated eggs. Mass culture of embryonated
 CC eggs is relatively cheap. NDV vaccines are relatively stable and can be
 CC simply administered by mass application methods e.g. drinking water or
 CC by spraying or by aerosol formation. The natural route of infection is
 CC by the respiratory and/or gastrointestinal tract which are also the major
 CC routes of infection of many other poultry pathogens. NDV can induce local
 CC immunity despite the presence of circulating maternal antibody. This
 CC sequence represents a protein encoded by the NDV strain LaSota genome
 CC which is described in the method of the invention.

XX Sequence 2204 AA:

SO

QY 1 FKPLMR 6
 :|||||
 Db 1844 frplmr 1849

Query Match 92.1%; Score 35; DB 21; Length 2204;
 Best Local Similarity 83.3%; Pred. No. 2.4e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

RESULT 3
 AAR30577
 ID AAR30577 standard; peptide: 6 AA.
 XX
 AC AAR30577:
 XX
 DT 29-JAN-1993 (first entry)
 XX
 DE Anaphylatoxin antagonist or agonist peptide.
 XX
 KW Anaphylatoxin: receptor; ligand; antiinflammatory; hexapeptide;
 KW heptapeptide; immunodeficiency; allergy; autoimmune; AIDS; cancer;
 KW infection; endotoxin; asthma; gout; psoriasis; cirrhosis;
 KW inflammatory; bowel; disease; hepatitis; burns; myocardial;
 KW infarction; transplant rejection; ischemic.
 XX
 OS Synthetic.
 XX
 FH Key
 FH Modified-site 1 Location/Qualifiers
 FT /note= "N-Me-Phe"
 FT Modified-site 4
 FT /label= OTHER
 FT /note= "(2R)-2-amino-3-cyclohexyl-propanoyl"
 FT Modified-site 6
 FT /note= "D-Arg"
 XX
 PN WO9211858-A.
 XX
 PD 23-JUL-1992.
 XX
 PF 10-DEC-1991; 91WO-US09319.
 XX
 PR 27-DEC-1990; 90US-0634641.
 XX
 PI (ABBO) ABBOTT LABORATORIES.
 XX
 PI Kawai M, Luly JR, Or YS, Wagner R, Wiedeman PE;
 XX
 DR WPI: 1992-268383/32.
 XX
 XX New hexa- and hepta-peptide(s) are anaphylatoxin antagonists and
 PT agonists - for treating inflammatory and immunodeficiency

PT diseases, cancers and severe infections
XX
PS Claim 11: Page 153; 160pp; English.
XX
CC The peptide is a specifically claimed example of a group of highly
CC generic hexa- and heptapeptides which are (a) anaphylatoxin
CC antagonists useful for treating asthma, other allergies,
CC inflammations, autoimmune diseases, serum sickness, gout, bullous
CC skin diseases, psoriasis, ARDS, endotoxin shock, hepatic cirrhosis,
CC pancreatitis, inflammatory bowel disease, burns, sepsis, myocardial
CC infarction, chronic hepatitis, transplant rejection, or ischaemic
CC heart or brain damage; or (b) anaphylatoxin agonists useful for
CC stimulating inflammatory and immune responses, e.g. in the treatment
CC of cancer, immunodeficiency diseases and severe infections.
XX
SQ Sequence 6 AA:

Query Match 86.8%; Score 33; DB 13; Length 6;
Best Local Similarity 83.3%; Pred. No. 4.3e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 FKPLMR 6
DB 1 fkpwxr 6

RESULT 4
AAR89737
ID AAR89737 standard; peptide: 6 AA.
XX
AC AAR89737:
XX
DT 03-MAY-1996 (first entry)
XX
DE C5a peptide analogue, C089, used to identify C5a (ant)agonists.
XX
KW C5a: complement; agonist; antagonist; peptide; binding assay;
KW Identification; inflammation; pain reduction; respiratory disorder;
KW cardiovascular; hypertension; ARDS; anaphylaxis; arthritis; asthma;
KW competitive.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 1 /label= Phe, OTHER
FT /note= "OTHER = the Bolton-Hunter modified peptide
FT having the 3-(p-hydroxyphenyl)-propionyl
FT group "
FT Misc-difference 4 /label= OTHER
FT /note= "D-cyclohexylalanine"
FT Modified-site 6 /label= OTHER
FT /note= "cyclohexylalanine"
XX
PN W09525957-A1.
XX
PD 28-SEP-1995.
XX
PE 14-MAR-1995; 95WO-US03209.
XX
PR 18-MAR-1994; 94US-0215137.
XX
PA (MERI) MERCK & CO INC.
XX
PI Konteacis Z, Siciliano SJ, Springer MS;
XX
DR WPL; 1995-344718/44.
XX
PT C5a receptor binding assays - used for identifying cpds having C5a
PT antagonist, agonist or partial agonist activity

XX
PS Claim 15: Page 47; 65pp; English.
XX
CC C5a C-terminal peptide analogues AAR89734-44 and AAR90033-35 may be used
CC in a new method for identifying C5a agonists/antagonists. AAR89734
CC represents the generic formula of this newly defined class of
CC labelled peptides. C5a antagonists identified using this method are
CC useful in the treatment of a wide variety of C5a-mediated diseases
CC such as acute respiratory distress syndrome (ARDS), anaphylactic
CC shock, psoriasis, osteoarthritis, rheumatoid arthritis, asthma,
CC cardiovascular disorders, and metastatic spread of cancerous tumours.
XX
SQ Sequence 6 AA:

Query Match 86.8%; Score 33; DB 16; Length 6;
Best Local Similarity 83.3%; Pred. No. 4.3e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 FKPLMR 6
DB 1 fkpwxr 6

RESULT 5
AAC10416
ID AAC10416 standard; Protein: 128 AA.
XX
AC AAC10416:
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 8729.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
FN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132046.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138040.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155119.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.

PR 26-OCT-1999: 99US-0161359.
PR 26-OCT-1999: 99US-0161360.
PR 26-OCT-1999: 99US-0161361.
PR 28-OCT-1999: 99US-0161920.
PR 28-OCT-1999: 99US-0161992.
PR 28-OCT-1999: 99US-0161993.
PR 29-OCT-1999: 99US-0162142.

Query Match 86.8%; Score 33; DB 21; Length 128;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1 Fkplw 5
Db 78 Fkplw 82

RESULT 6
AAC52516
ID AAC52516 standard; Protein: 128 AA.
XX
AC AAC52516:
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 66766.
XX
KW Protein Identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000: 2000EP-0301439.
XX
PF 25-FEB-1999: 99US-0121825.
XX 05-MAR-1999: 99US-0123180.
PR 09-MAR-1999: 99US-0123548.
PR 23-MAR-1999: 99US-0125788.
PR 25-MAR-1999: 99US-0126264.
PR 29-MAR-1999: 99US-0126785.
PR 01-APR-1999: 99US-0127462.
PR 06-APR-1999: 99US-0128234.
PR 08-APR-1999: 99US-0128714.
PR 16-APR-1999: 99US-0129845.
PR 19-APR-1999: 99US-0130077.
PR 21-APR-1999: 99US-0130449.
PR 23-APR-1999: 99US-0130510.
PR 23-APR-1999: 99US-0130891.
PR 28-APR-1999: 99US-0131449.
PR 30-APR-1999: 99US-0132048.
PR 30-APR-1999: 99US-0132407.
PR 04-MAY-1999: 99US-0132484.
PR 05-MAY-1999: 99US-0132485.
PR 06-MAY-1999: 99US-0132486.
PR 06-MAY-1999: 99US-0132487.
PR 07-MAY-1999: 99US-0132863.
PR 11-MAY-1999: 99US-0134256.
PR 14-MAY-1999: 99US-0134218.
PR 14-MAY-1999: 99US-0134219.
PR 14-MAY-1999: 99US-0134221.
PR 14-MAY-1999: 99US-0134370.
PR 18-MAY-1999: 99US-0134768.
PR 19-MAY-1999: 99US-0134941.
PR 20-MAY-1999: 99US-0135124.
PR 21-MAY-1999: 99US-0135353.
PR 24-MAY-1999: 99US-0135629.
PR 25-MAY-1999: 99US-0136021.
PR 27-MAY-1999: 99US-0136392.

PR 28-MAY-1999: 99US-0136782.
PR 01-JUN-1999: 99US-0137222.
PR 03-JUN-1999: 99US-0137528.
PR 04-JUN-1999: 99US-0137502.
PR 07-JUN-1999: 99US-0137724.
PR 08-JUN-1999: 99US-0138094.
PR 10-JUN-1999: 99US-0138540.
PR 10-JUN-1999: 99US-0138847.
PR 14-JUN-1999: 99US-0139119.
PR 16-JUN-1999: 99US-0139452.
PR 16-JUN-1999: 99US-0139453.
PR 17-JUN-1999: 99US-0139492.
PR 18-JUN-1999: 99US-0139454.
PR 18-JUN-1999: 99US-0139455.
PR 18-JUN-1999: 99US-0139456.
PR 18-JUN-1999: 99US-0139457.
PR 18-JUN-1999: 99US-0139458.
PR 18-JUN-1999: 99US-0139459.
PR 18-JUN-1999: 99US-0139460.
PR 18-JUN-1999: 99US-0139461.
PR 18-JUN-1999: 99US-0139462.
PR 18-JUN-1999: 99US-0139463.
PR 18-JUN-1999: 99US-0139750.
PR 18-JUN-1999: 99US-0139763.
PR 21-JUN-1999: 99US-0139817.
PR 22-JUN-1999: 99US-0139899.
PR 23-JUN-1999: 99US-0140353.
PR 23-JUN-1999: 99US-0140354.
PR 24-JUN-1999: 99US-0140695.
PR 28-JUN-1999: 99US-0140823.
PR 29-JUN-1999: 99US-0140991.
PR 30-JUN-1999: 99US-0141287.
PR 01-JUL-1999: 99US-0141842.
PR 01-JUL-1999: 99US-0142154.
PR 02-JUL-1999: 99US-0142055.
PR 06-JUL-1999: 99US-0142390.
PR 08-JUL-1999: 99US-0142803.
PR 09-JUL-1999: 99US-0142920.
PR 12-JUL-1999: 99US-0142977.
PR 13-JUL-1999: 99US-0143542.
PR 14-JUL-1999: 99US-0143624.
PR 15-JUL-1999: 99US-0144005.
PR 16-JUL-1999: 99US-0144085.
PR 16-JUL-1999: 99US-0144086.
PR 19-JUL-1999: 99US-0144325.
PR 19-JUL-1999: 99US-0144331.
PR 19-JUL-1999: 99US-0144332.
PR 19-JUL-1999: 99US-0144333.
PR 19-JUL-1999: 99US-0144334.
PR 19-JUL-1999: 99US-0144335.
PR 20-JUL-1999: 99US-0144352.
PR 20-JUL-1999: 99US-0144632.
PR 20-JUL-1999: 99US-0144684.
PR 21-JUL-1999: 99US-0144814.
PR 21-JUL-1999: 99US-0145086.
PR 21-JUL-1999: 99US-0145088.
PR 22-JUL-1999: 99US-0145087.
PR 22-JUL-1999: 99US-0145089.
PR 22-JUL-1999: 99US-0145192.
PR 22-JUL-1999: 99US-0145192.
PR 23-JUL-1999: 99US-0145145.
PR 23-JUL-1999: 99US-0145218.
PR 23-JUL-1999: 99US-0145224.
PR 26-JUL-1999: 99US-0145276.
PR 27-JUL-1999: 99US-0145913.
PR 27-JUL-1999: 99US-0145918.
PR 27-JUL-1999: 99US-0145919.
PR 28-JUL-1999: 99US-0145951.
PR 02-AUG-1999: 99US-0146386.
PR 02-AUG-1999: 99US-0146388.
PR 02-AUG-1999: 99US-0146389.
PR 03-AUG-1999: 99US-0147038.
PR 04-AUG-1999: 99US-0147204.

PR 04-AUG-1999: 99US-0147302.
PR 05-AUG-1999: 99US-0147192.
PR 05-AUG-1999: 99US-0147260.
PR 06-AUG-1999: 99US-0147203.
PR 06-AUG-1999: 99US-0147416.
PR 09-AUG-1999: 99US-0147493.
PR 09-AUG-1999: 99US-0147935.
PR 10-AUG-1999: 99US-0148171.
PR 11-AUG-1999: 99US-0148319.
PR 12-AUG-1999: 99US-0148341.
PR 13-AUG-1999: 99US-0148565.
PR 13-AUG-1999: 99US-0148684.
PR 16-AUG-1999: 99US-0149368.
PR 17-AUG-1999: 99US-0149175.
PR 18-AUG-1999: 99US-0149426.
PR 20-AUG-1999: 99US-0149722.
PR 20-AUG-1999: 99US-0149723.
PR 20-AUG-1999: 99US-0149929.
PR 23-AUG-1999: 99US-0149902.
PR 23-AUG-1999: 99US-0149930.
PR 25-AUG-1999: 99US-0150566.
PR 26-AUG-1999: 99US-0150884.
PR 27-AUG-1999: 99US-0151065.
PR 27-AUG-1999: 99US-0151066.
PR 27-AUG-1999: 99US-0151080.
PR 30-AUG-1999: 99US-0151303.
PR 31-AUG-1999: 99US-0151438.
PR 01-SEP-1999: 99US-0151930.
PR 07-SEP-1999: 99US-0152363.
PR 10-SEP-1999: 99US-0153070.
PR 13-SEP-1999: 99US-0153758.
PR 15-SEP-1999: 99US-0154018.
PR 16-SEP-1999: 99US-0154039.
PR 20-SEP-1999: 99US-0154779.
PR 22-SEP-1999: 99US-0155139.
PR 23-SEP-1999: 99US-0155486.
PR 24-SEP-1999: 99US-0155659.
PR 28-SEP-1999: 99US-0156458.
PR 29-SEP-1999: 99US-0156596.
PR 04-OCT-1999: 99US-0157117.
PR 05-OCT-1999: 99US-0157753.
PR 06-OCT-1999: 99US-0157865.
PR 07-OCT-1999: 99US-0158029.
PR 08-OCT-1999: 99US-0158332.
PR 12-OCT-1999: 99US-0158369.
PR 13-OCT-1999: 99US-0159293.
PR 13-OCT-1999: 99US-0159294.
PR 13-OCT-1999: 99US-0159295.
PR 14-OCT-1999: 99US-0159329.
PR 14-OCT-1999: 99US-0159330.
PR 14-OCT-1999: 99US-0159331.
PR 14-OCT-1999: 99US-0159637.
PR 14-OCT-1999: 99US-0159638.
PR 18-OCT-1999: 99US-0159584.
PR 21-OCT-1999: 99US-0160741.
PR 21-OCT-1999: 99US-0160767.
PR 21-OCT-1999: 99US-0160768.
PR 21-OCT-1999: 99US-0160770.
PR 21-OCT-1999: 99US-0160814.
PR 21-OCT-1999: 99US-0160815.
PR 22-OCT-1999: 99US-0160980.
PR 22-OCT-1999: 99US-0160981.
PR 22-OCT-1999: 99US-0160989.
PR 25-OCT-1999: 99US-0161404.
PR 25-OCT-1999: 99US-0161405.
PR 25-OCT-1999: 99US-0161406.
PR 26-OCT-1999: 99US-0161359.
PR 26-OCT-1999: 99US-0161360.
PR 26-OCT-1999: 99US-0161361.
PR 28-OCT-1999: 99US-0161920.
PR 28-OCT-1999: 99US-0161992.
PR 28-OCT-1999: 99US-0161993.
PR 29-OCT-1999: 99US-0162142.

Query Match 86.8%; Score 33; DB 21; Length 128;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EKPLW 5
Db 78 EKPLW 82
RESULT 7
AAW74783
ID AAW74783 standard; Protein: 152 AA.
XX
AC AAW74783;
XX
UT 25-JAN-1999 (first entry)
XX
XX Human secreted protein encoded by gene 54 clone HMADJ02.
XX
XX Human: secreted protein; testis; tumour; foetal brain tissue;
KW fusion protein; cancer; central nervous system; seizure;
KW diagnosis; neurodegenerative disease.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 146 label= unknown
FT Misc-difference 152 label= unknown
FT
XX
PN WO9839448-A2.
XX
PD 11-SEP-1998.
XX
PF 06-MAR-1998: 98WO-US04493.
XX
PR 02-OCT-1997: 97US-0061060.
PR 07-MAR-1997: 97US-0038621.
PR 07-MAR-1997: 97US-0040161.
PR 07-MAR-1997: 97US-0040162.
PR 07-MAR-1997: 97US-0040163.
PR 07-MAR-1997: 97US-0040333.
PR 07-MAR-1997: 97US-0040334.
PR 07-MAR-1997: 97US-0040336.
PR 07-MAR-1997: 97US-0040626.
PR 11-APR-1997: 97US-0043311.
PR 11-APR-1997: 97US-0043312.
PR 11-APR-1997: 97US-0043313.
PR 11-APR-1997: 97US-0043314.
PR 11-APR-1997: 97US-0043568.
PR 11-APR-1997: 97US-0043569.
PR 11-APR-1997: 97US-0043576.
PR 11-APR-1997: 97US-0043578.
PR 11-APR-1997: 97US-0043580.
PR 11-APR-1997: 97US-0043659.
PR 11-APR-1997: 97US-0043670.
PR 11-APR-1997: 97US-0043671.
PR 11-APR-1997: 97US-0043672.
PR 11-APR-1997: 97US-0043674.
PR 23-MAY-1997: 97US-0047492.
PR 23-MAY-1997: 97US-0047500.
PR 23-MAY-1997: 97US-0047501.
PR 23-MAY-1997: 97US-0047502.
PR 23-MAY-1997: 97US-0047503.
PR 23-MAY-1997: 97US-0047581.
PR 23-MAY-1997: 97US-0047582.
PR 23-MAY-1997: 97US-0047583.
PR 23-MAY-1997: 97US-0047584.
PR 23-MAY-1997: 97US-0047585.
PR 23-MAY-1997: 97US-0047586.
PR 23-MAY-1997: 97US-0047587.

PR 23-MAY-1997; 97US-0047588.
PR 23-MAY-1997; 97US-0047589.
PR 23-MAY-1997; 97US-0047590.
PR 23-MAY-1997; 97US-0047592.
PR 23-MAY-1997; 97US-0047593.
PR 23-MAY-1997; 97US-0047594.
PR 23-MAY-1997; 97US-0047595.
PR 23-MAY-1997; 97US-0047596.
PR 23-MAY-1997; 97US-0047597.
PR 23-MAY-1997; 97US-0047598.
PR 23-MAY-1997; 97US-0047599.
PR 23-MAY-1997; 97US-0047600.
PR 23-MAY-1997; 97US-0047601.
PR 23-MAY-1997; 97US-0047612.
PR 23-MAY-1997; 97US-0047613.
PR 23-MAY-1997; 97US-0047614.
PR 23-MAY-1997; 97US-0047615.
PR 23-MAY-1997; 97US-0047617.
PR 23-MAY-1997; 97US-0047618.
PR 23-MAY-1997; 97US-0047632.
PR 23-MAY-1997; 97US-0047633.
PR 06-JUN-1997; 97US-0048964.
PR 06-JUN-1997; 97US-0048974.
PR 13-JUN-1997; 97US-0049610.
PR 08-JUL-1997; 97US-0051926.
PR 16-JUL-1997; 97US-0052874.
PR 18-AUG-1997; 97US-0055724.
PR 22-AUG-1997; 97US-0056630.
PR 22-AUG-1997; 97US-0056631.
PR 22-AUG-1997; 97US-0056632.
PR 22-AUG-1997; 97US-0056636.
PR 22-AUG-1997; 97US-0056637.
PR 22-AUG-1997; 97US-0056664.
PR 22-AUG-1997; 97US-0056845.
PR 22-AUG-1997; 97US-0056862.
PR 22-AUG-1997; 97US-0056864.
PR 22-AUG-1997; 97US-0056872.
PR 22-AUG-1997; 97US-0056874.
PR 22-AUG-1997; 97US-0056875.
PR 22-AUG-1997; 97US-0056876.
PR 22-AUG-1997; 97US-0056877.
PR 22-AUG-1997; 97US-0056878.
PR 22-AUG-1997; 97US-0056879.
PR 22-AUG-1997; 97US-0056880.
PR 22-AUG-1997; 97US-0056881.
PR 22-AUG-1997; 97US-0056882.
PR 22-AUG-1997; 97US-0056884.
PR 22-AUG-1997; 97US-0056886.
PR 22-AUG-1997; 97US-0056887.
PR 22-AUG-1997; 97US-0056888.
PR 22-AUG-1997; 97US-0056889.
PR 22-AUG-1997; 97US-0056892.
PR 22-AUG-1997; 97US-0056893.
PR 22-AUG-1997; 97US-0056894.
PR 22-AUG-1997; 97US-0056903.
PR 22-AUG-1997; 97US-0056908.
PR 22-AUG-1997; 97US-0056909.
PR 22-AUG-1997; 97US-0056910.
PR 22-AUG-1997; 97US-0056911.
PR 05-SEP-1997; 97US-0057650.
PR 05-SEP-1997; 97US-0057659.
PR 05-SEP-1997; 97US-0057661.
PR 12-SEP-1997; 97US-0058785.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress CA;
PI Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS;
PI Kyaw H, Lallieur DM, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu CL, Zeng Z;
DR PTL: 1998-506364/43.

DR N-PSDB: AAV59564.
XX
XX New isolated human genes and the secreted polypeptide(s) they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
XX
XX
PS Claim 1: Page 572-573; 721pp; English.
XX
XX This sequence represents a secreted human protein encoded by the nucleic
CC acid molecule designated Gene 54 from the human cDNA clone HMADJ02
CC (deposited as clone ATCC 97899 and ATCC 2090451).
CC The gene can be used to generate fusion proteins by linking to the gene
CC to a human immunoglobulin Fc portion (e.g. AAV59502) for increasing the
CC stability of the fused protein as compared to the human protein only.
CC The invention relates to 186 novel genes and their fragments (nucleic
CC acid sequences: AAV59511-V59812; amino acid sequences AAV74731-W75026)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 186
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AAV59511 for described uses).
XX
SQ Sequence 152 AA:

Query Match 86.8%; Score 33; DB 19; Length 152;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FRPLW 5
11111
Db 121 fklpw 125

RESULT 8
AAAG10415
ID AAAG10415 standard; Protein: 174 AA.
XX
XX AAAG10415;
XX
XX 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 8728.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999: 99US-0132048.
PR 30-APR-1999: 99US-0132407.
PR 04-MAY-1999: 99US-0132484.
PR 05-MAY-1999: 99US-0132485.
PR 06-MAY-1999: 99US-0132486.
PR 06-MAY-1999: 99US-0132487.
PR 07-MAY-1999: 99US-0132487.
PR 11-MAY-1999: 99US-0134256.
PR 14-MAY-1999: 99US-0134218.
PR 14-MAY-1999: 99US-0134219.
PR 14-MAY-1999: 99US-0134221.
PR 14-MAY-1999: 99US-0134370.
PR 18-MAY-1999: 99US-0134768.
PR 19-MAY-1999: 99US-0134941.
PR 20-MAY-1999: 99US-0135124.
PR 21-MAY-1999: 99US-0135353.
PR 24-MAY-1999: 99US-0135629.
PR 25-MAY-1999: 99US-0136021.
PR 27-MAY-1999: 99US-0136392.
PR 28-MAY-1999: 99US-0136782.
PR 01-JUN-1999: 99US-0137222.
PR 03-JUN-1999: 99US-0137528.
PR 04-JUN-1999: 99US-0137502.
PR 07-JUN-1999: 99US-0137724.
PR 08-JUN-1999: 99US-0138094.
PR 10-JUN-1999: 99US-0138540.
PR 10-JUN-1999: 99US-0138847.
PR 14-JUN-1999: 99US-0139119.
PR 16-JUN-1999: 99US-0139452.
PR 16-JUN-1999: 99US-0139453.
PR 17-JUN-1999: 99US-0139492.
PR 18-JUN-1999: 99US-0139454.
PR 18-JUN-1999: 99US-0139455.
PR 18-JUN-1999: 99US-0139456.
PR 18-JUN-1999: 99US-0139457.
PR 18-JUN-1999: 99US-0139458.
PR 18-JUN-1999: 99US-0139459.
PR 18-JUN-1999: 99US-0139460.
PR 18-JUN-1999: 99US-0139461.
PR 18-JUN-1999: 99US-0139462.
PR 18-JUN-1999: 99US-0139463.
PR 18-JUN-1999: 99US-0139750.
PR 18-JUN-1999: 99US-0139763.
PR 21-JUN-1999: 99US-0139817.
PR 22-JUN-1999: 99US-0139899.
PR 23-JUN-1999: 99US-0140353.
PR 23-JUN-1999: 99US-0140354.
PR 24-JUN-1999: 99US-0140695.
PR 28-JUN-1999: 99US-0140823.
PR 29-JUN-1999: 99US-0140991.
PR 30-JUN-1999: 99US-0141287.
PR 01-JUL-1999: 99US-0141842.
PR 01-JUL-1999: 99US-0142154.
PR 02-JUL-1999: 99US-0142055.
PR 06-JUL-1999: 99US-0142390.
PR 08-JUL-1999: 99US-0142803.
PR 09-JUL-1999: 99US-0142920.
PR 12-JUL-1999: 99US-0142927.
PR 13-JUL-1999: 99US-0143542.
PR 14-JUL-1999: 99US-0143624.
PR 15-JUL-1999: 99US-0144005.
PR 16-JUL-1999: 99US-0144085.
PR 16-JUL-1999: 99US-0144086.
PR 19-JUL-1999: 99US-0144325.
PR 19-JUL-1999: 99US-0144331.
PR 19-JUL-1999: 99US-0144332.
PR 19-JUL-1999: 99US-0144333.
PR 19-JUL-1999: 99US-0144334.
PR 19-JUL-1999: 99US-0144335.
PR 20-JUL-1999: 99US-0144352.
PR 20-JUL-1999: 99US-0144632.
PR 20-JUL-1999: 99US-0144884.
PR 21-JUL-1999: 99US-0144814.

PR 21-JUL-1999: 99US-0145086.
PR 21-JUL-1999: 99US-0145088.
PR 22-JUL-1999: 99US-0145085.
PR 22-JUL-1999: 99US-0145087.
PR 22-JUL-1999: 99US-0145089.
PR 22-JUL-1999: 99US-0145192.
PR 23-JUL-1999: 99US-0145145.
PR 23-JUL-1999: 99US-0145224.
PR 23-JUL-1999: 99US-0145228.
PR 26-JUL-1999: 99US-0145276.
PR 27-JUL-1999: 99US-0145913.
PR 27-JUL-1999: 99US-0145918.
PR 27-JUL-1999: 99US-0145919.
PR 28-JUL-1999: 99US-0145951.
PR 02-AUG-1999: 99US-0146386.
PR 02-AUG-1999: 99US-0146388.
PR 02-AUG-1999: 99US-0146389.
PR 03-AUG-1999: 99US-0147038.
PR 04-AUG-1999: 99US-0147204.
PR 04-AUG-1999: 99US-0147302.
PR 05-AUG-1999: 99US-0147192.
PR 05-AUG-1999: 99US-0147260.
PR 06-AUG-1999: 99US-0147303.
PR 06-AUG-1999: 99US-0147416.
PR 09-AUG-1999: 99US-0147493.
PR 10-AUG-1999: 99US-0147935.
PR 11-AUG-1999: 99US-0148171.
PR 12-AUG-1999: 99US-0148319.
PR 13-AUG-1999: 99US-0148341.
PR 13-AUG-1999: 99US-0148565.
PR 16-AUG-1999: 99US-0148684.
PR 17-AUG-1999: 99US-0149175.
PR 18-AUG-1999: 99US-0149426.
PR 20-AUG-1999: 99US-0149722.
PR 20-AUG-1999: 99US-0149723.
PR 20-AUG-1999: 99US-0149929.
PR 23-AUG-1999: 99US-0149902.
PR 23-AUG-1999: 99US-0149930.
PR 25-AUG-1999: 99US-0150566.
PR 26-AUG-1999: 99US-0150884.
PR 27-AUG-1999: 99US-0151065.
PR 27-AUG-1999: 99US-0151066.
PR 30-AUG-1999: 99US-0151080.
PR 30-AUG-1999: 99US-0151303.
PR 31-AUG-1999: 99US-0151438.
PR 01-SEP-1999: 99US-0151930.
PR 07-SEP-1999: 99US-0152363.
PR 10-SEP-1999: 99US-0153070.
PR 13-SEP-1999: 99US-0153758.
PR 15-SEP-1999: 99US-0154018.
PR 16-SEP-1999: 99US-0154039.
PR 20-SEP-1999: 99US-0154779.
PR 22-SEP-1999: 99US-0155139.
PR 23-SEP-1999: 99US-0155486.
PR 24-SEP-1999: 99US-0155659.
PR 28-SEP-1999: 99US-0156458.
PR 29-SEP-1999: 99US-0156596.
PR 04-OCT-1999: 99US-0157117.
PR 05-OCT-1999: 99US-0157753.
PR 06-OCT-1999: 99US-0157865.
PR 07-OCT-1999: 99US-0158029.
PR 08-OCT-1999: 99US-0158232.
PR 12-OCT-1999: 99US-0158369.
PR 13-OCT-1999: 99US-0159293.
PR 13-OCT-1999: 99US-0159294.
PR 13-OCT-1999: 99US-0159295.
PR 14-OCT-1999: 99US-0159329.
PR 14-OCT-1999: 99US-0159330.
PR 14-OCT-1999: 99US-0159331.
PR 14-OCT-1999: 99US-0159637.
PR 14-OCT-1999: 99US-0159638.
PR 18-OCT-1999: 99US-0159584.

PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161982.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 86.8%; Score 33; DB 21; Length 174;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Caps 0;

OY 1 FKPLW 5
| | | | |
Db 124 fkp1w 128

RESULT 9

AAG52515
ID AAG52515 standard; Protein: 174 AA.

AC AAG52515;

DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 66765.

XX Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.

XX Arabidopsis thaliana.

PN EPI033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999: 99US-0145218.
PR 23-JUL-1999: 99US-0145224.
PR 26-JUL-1999: 99US-0145276.
PR 27-JUL-1999: 99US-0145913.
PR 27-JUL-1999: 99US-0145918.
PR 27-JUL-1999: 99US-0145919.
PR 28-JUL-1999: 99US-0145951.
PR 02-AUG-1999: 99US-0146386.
PR 02-AUG-1999: 99US-0146388.
PR 02-AUG-1999: 99US-0146389.
PR 04-AUG-1999: 99US-0147038.
PR 04-AUG-1999: 99US-0147204.
PR 05-AUG-1999: 99US-0147302.
PR 05-AUG-1999: 99US-0147302.
PR 05-AUG-1999: 99US-0147302.
PR 06-AUG-1999: 99US-0147303.
PR 06-AUG-1999: 99US-0147303.
PR 06-AUG-1999: 99US-0147416.
PR 09-AUG-1999: 99US-0147493.
PR 09-AUG-1999: 99US-0147935.
PR 10-AUG-1999: 99US-0148171.
PR 11-AUG-1999: 99US-0148319.
PR 12-AUG-1999: 99US-0148341.
PR 13-AUG-1999: 99US-0148565.
PR 13-AUG-1999: 99US-0148684.
PR 16-AUG-1999: 99US-0149368.
PR 17-AUG-1999: 99US-0149375.
PR 18-AUG-1999: 99US-0149426.
PR 20-AUG-1999: 99US-0149722.
PR 20-AUG-1999: 99US-0149723.
PR 20-AUG-1999: 99US-0149723.
PR 20-AUG-1999: 99US-0149929.
PR 23-AUG-1999: 99US-0149902.
PR 23-AUG-1999: 99US-0149930.
PR 25-AUG-1999: 99US-0150566.
PR 26-AUG-1999: 99US-0150884.
PR 27-AUG-1999: 99US-0151065.
PR 27-AUG-1999: 99US-0151066.
PR 27-AUG-1999: 99US-0151080.
PR 30-AUG-1999: 99US-0151303.
PR 31-AUG-1999: 99US-0151438.
PR 01-SEP-1999: 99US-0151930.
PR 07-SEP-1999: 99US-0152363.
PR 10-SEP-1999: 99US-0153070.
PR 13-SEP-1999: 99US-0153758.
PR 15-SEP-1999: 99US-0154018.
PR 16-SEP-1999: 99US-0154039.
PR 20-SEP-1999: 99US-0154779.
PR 22-SEP-1999: 99US-0155139.
PR 23-SEP-1999: 99US-0155486.
PR 24-SEP-1999: 99US-0155659.
PR 28-SEP-1999: 99US-0156458.
PR 29-SEP-1999: 99US-0156596.
PR 04-OCT-1999: 99US-0157117.
PR 05-OCT-1999: 99US-0157753.
PR 06-OCT-1999: 99US-0157865.
PR 07-OCT-1999: 99US-0158029.
PR 08-OCT-1999: 99US-0158232.
PR 12-OCT-1999: 99US-0158369.
PR 13-OCT-1999: 99US-0159293.
PR 13-OCT-1999: 99US-0159294.
PR 13-OCT-1999: 99US-0159295.
PR 14-OCT-1999: 99US-0159329.
PR 14-OCT-1999: 99US-0159330.
PR 14-OCT-1999: 99US-0159331.
PR 14-OCT-1999: 99US-0159637.
PR 14-OCT-1999: 99US-0159638.
PR 18-OCT-1999: 99US-0159584.
PR 21-OCT-1999: 99US-0160741.
PR 21-OCT-1999: 99US-0160767.
PR 21-OCT-1999: 99US-0160768.
PR 21-OCT-1999: 99US-0160770.
PR 21-OCT-1999: 99US-0160814.
PR 21-OCT-1999: 99US-0160815.
PR 22-OCT-1999: 99US-0160980.

FR 22-OCT-1999: 99US-0160981.
FR 22-OCT-1999: 99US-0160989.
PR 25-OCT-1999: 99US-0161404.
PR 25-OCT-1999: 99US-0161405.
PR 25-OCT-1999: 99US-0161406.
PR 26-OCT-1999: 99US-0161359.
PR 26-OCT-1999: 99US-0161360.
PR 26-OCT-1999: 99US-0161361.
PR 28-OCT-1999: 99US-0161920.
PR 28-OCT-1999: 99US-0161992.
PR 28-OCT-1999: 99US-0161993.
PR 29-OCT-1999: 99US-0162142.

Query Match 86.8%; Score 33; DB 21; Length 174;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1 FKPLW 5
DB 124 FKPLW 128

RESULT 10
AAC33129
ID AAC33129 standard; Protein: 182 AA.
XX
AC AAC33129;
XX
DT 18-OCT-2000 (first entry)
XX
DE Zea mays protein fragment SEQ ID NO: 40095.
XX
KW Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.
XX
OS Zea mays subsp. mays.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX
FR 25-FEB-1999: 99US-0121825.
FR 05-MAR-1999: 99US-0123180.
PR 09-MAR-1999: 99US-0123548.
PR 23-MAR-1999: 99US-0125788.
PR 25-MAR-1999: 99US-0126264.
PR 28-MAR-1999: 99US-0126785.
PR 01-APR-1999: 99US-0127462.
PR 06-APR-1999: 99US-0128234.
PR 08-APR-1999: 99US-0128714.
PR 16-APR-1999: 99US-0129845.
PR 19-APR-1999: 99US-0130077.
PR 21-APR-1999: 99US-0130449.
PR 23-APR-1999: 99US-0130510.
PR 23-APR-1999: 99US-0130891.
PR 28-APR-1999: 99US-0131449.
PR 30-APR-1999: 99US-0132048.
PR 30-APR-1999: 99US-0132407.
PR 04-MAY-1999: 99US-0132484.
PR 05-MAY-1999: 99US-0132485.
PR 06-MAY-1999: 99US-0132486.
PR 06-MAY-1999: 99US-0132487.
PR 07-MAY-1999: 99US-0132863.
PR 11-MAY-1999: 99US-0134256.
PR 14-MAY-1999: 99US-0134218.
PR 14-MAY-1999: 99US-0134219.
PR 14-MAY-1999: 99US-0134421.
PR 14-MAY-1999: 99US-0134470.
PR 18-MAY-1999: 99US-0134768.
PR 19-MAY-1999: 99US-0134941.

PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.

FR 02-AUG-1999; 99US-0146386.
FR 02-AUG-1999; 99US-0146388.
FR 02-AUG-1999; 99US-0146389.
FR 03-AUG-1999; 99US-0147038.
FR 04-AUG-1999; 99US-0147204.
FR 04-AUG-1999; 99US-0147302.
FR 05-AUG-1999; 99US-0147192.
FR 05-AUG-1999; 99US-0147260.
FR 06-AUG-1999; 99US-0147303.
FR 06-AUG-1999; 99US-0147416.
FR 09-AUG-1999; 99US-0147493.
FR 09-AUG-1999; 99US-0147935.
FR 10-AUG-1999; 99US-0148171.
FR 11-AUG-1999; 99US-0148319.
FR 12-AUG-1999; 99US-0148341.
FR 13-AUG-1999; 99US-0148565.
FR 13-AUG-1999; 99US-0148684.
FR 16-AUG-1999; 99US-0149175.
FR 17-AUG-1999; 99US-0149368.
FR 18-AUG-1999; 99US-0149426.
FR 20-AUG-1999; 99US-0149722.
FR 20-AUG-1999; 99US-0149723.
FR 20-AUG-1999; 99US-0149929.
FR 23-AUG-1999; 99US-0149902.
FR 23-AUG-1999; 99US-0149930.
FR 25-AUG-1999; 99US-0150566.
FR 26-AUG-1999; 99US-0150884.
FR 27-AUG-1999; 99US-0151065.
FR 27-AUG-1999; 99US-0151066.
FR 27-AUG-1999; 99US-0151080.
FR 30-AUG-1999; 99US-0151303.
FR 31-AUG-1999; 99US-0151438.
FR 01-SEP-1999; 99US-0151930.
FR 07-SEP-1999; 99US-0152363.
FR 10-SEP-1999; 99US-0153070.
FR 13-SEP-1999; 99US-0153758.
FR 15-SEP-1999; 99US-0154018.
FR 16-SEP-1999; 99US-0154039.
FR 20-SEP-1999; 99US-0154779.
FR 22-SEP-1999; 99US-0155139.
FR 23-SEP-1999; 99US-0155486.
FR 24-SEP-1999; 99US-0155659.
FR 28-SEP-1999; 99US-0156458.
FR 29-SEP-1999; 99US-0156596.
FR 04-OCT-1999; 99US-0157117.
FR 05-OCT-1999; 99US-0157753.
FR 06-OCT-1999; 99US-0157865.
FR 07-OCT-1999; 99US-0158029.
FR 08-OCT-1999; 99US-0158232.
FR 12-OCT-1999; 99US-0158369.
FR 13-OCT-1999; 99US-0159293.
FR 13-OCT-1999; 99US-0159294.
FR 13-OCT-1999; 99US-0159295.
FR 14-OCT-1999; 99US-0159329.
FR 14-OCT-1999; 99US-0159330.
FR 14-OCT-1999; 99US-0159331.
FR 14-OCT-1999; 99US-0159637.
FR 14-OCT-1999; 99US-0159638.
FR 18-OCT-1999; 99US-0159584.
FR 21-OCT-1999; 99US-0160741.
FR 21-OCT-1999; 99US-0160767.
FR 21-OCT-1999; 99US-0160768.
FR 21-OCT-1999; 99US-0160770.
FR 21-OCT-1999; 99US-0160814.
FR 21-OCT-1999; 99US-0160815.
FR 22-OCT-1999; 99US-0160980.
FR 22-OCT-1999; 99US-0160981.
FR 22-OCT-1999; 99US-0160989.
FR 25-OCT-1999; 99US-0161404.
FR 25-OCT-1999; 99US-0161405.
FR 25-OCT-1999; 99US-0161406.
FR 26-OCT-1999; 99US-0161359.
FR 26-OCT-1999; 99US-0161360.

PR 26-OCT-1999: 99US-0161361.
PR 28-OCT-1999: 99US-0161920.
PR 28-OCT-1999: 99US-0161992.
PR 28-OCT-1999: 99US-0161993.
PR 29-OCT-1999: 99US-0162142.

Query Match 86.8% Score 33: DB 21: Length 182:
Best Local Similarity 100.0% Pred. No. 45:
Matches 5: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

OY 1 EXPLW 5
Db 124 fXpLw 128

RESULT 11

AAG15288
ID AAG15288 standard; Protein: 257 AA.

AC AAG15288;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 15484.

XX Protein identification: signal transduction pathway: metabolic pathway:
KM hybridisation assay; genetic mapping; gene expression control; promoter:
KM termination sequence.

XX Arabidopsis thaliana.

PN EP103405-A2.

XX 06-SEP-2000.

PD 25-FEB-2000: 2000EP-0301439.

XX 25-FEB-1999: 99US-0121825.

PR 05-MAR-1999: 99US-0123180.

PR 09-MAR-1999: 99US-0123548.

PR 23-MAR-1999: 99US-0125788.

PR 25-MAR-1999: 99US-0126264.

PR 29-MAR-1999: 99US-0126785.

PR 01-APR-1999: 99US-0127462.

PR 06-APR-1999: 99US-0128234.

PR 08-APR-1999: 99US-0128714.

PR 16-APR-1999: 99US-0129845.

PR 19-APR-1999: 99US-0130077.

PR 21-APR-1999: 99US-0130449.

PR 23-APR-1999: 99US-0130510.

PR 28-APR-1999: 99US-0130891.

PR 30-APR-1999: 99US-0131449.

PR 30-APR-1999: 99US-0132048.

PR 04-MAY-1999: 99US-0132407.

PR 04-MAY-1999: 99US-0132484.

PR 05-MAY-1999: 99US-0132485.

PR 06-MAY-1999: 99US-0132486.

PR 06-MAY-1999: 99US-0132487.

PR 07-MAY-1999: 99US-0132863.

PR 11-MAY-1999: 99US-0134256.

PR 14-MAY-1999: 99US-0134218.

PR 14-MAY-1999: 99US-0134219.

PR 14-MAY-1999: 99US-0134321.

PR 14-MAY-1999: 99US-0134370.

PR 18-MAY-1999: 99US-0134768.

PR 19-MAY-1999: 99US-0134941.

PR 20-MAY-1999: 99US-0135124.

PR 21-MAY-1999: 99US-0135353.

PR 24-MAY-1999: 99US-0135629.

PR 25-MAY-1999: 99US-0136021.

PR 27-MAY-1999: 99US-0136392.

PR 28-MAY-1999: 99US-0136782.

PR 01-JUN-1999: 99US-0137222.

PR 03-JUN-1999: 99US-0137528.

PR 04-JUN-1999: 99US-0137502.

PR 07-JUN-1999: 99US-0137724.

PR 08-JUN-1999: 99US-0138094.

PR 10-JUN-1999: 99US-0138540.

PR 14-JUN-1999: 99US-0138847.

PR 16-JUN-1999: 99US-0139119.

PR 16-JUN-1999: 99US-0139452.

PR 17-JUN-1999: 99US-0139453.

PR 18-JUN-1999: 99US-0139492.

PR 18-JUN-1999: 99US-0139454.

PR 18-JUN-1999: 99US-0139455.

PR 18-JUN-1999: 99US-0139456.

PR 18-JUN-1999: 99US-0139457.

PR 18-JUN-1999: 99US-0139458.

PR 18-JUN-1999: 99US-0139459.

PR 18-JUN-1999: 99US-0139460.

PR 18-JUN-1999: 99US-0139461.

PR 18-JUN-1999: 99US-0139462.

PR 18-JUN-1999: 99US-0139463.

PR 18-JUN-1999: 99US-0139750.

PR 18-JUN-1999: 99US-0139763.

PR 21-JUN-1999: 99US-0139817.

PR 22-JUN-1999: 99US-0139899.

PR 23-JUN-1999: 99US-0140353.

PR 24-JUN-1999: 99US-0140354.

PR 28-JUN-1999: 99US-0140695.

PR 29-JUN-1999: 99US-0140823.

PR 30-JUN-1999: 99US-0140991.

PR 01-JUL-1999: 99US-0141287.

PR 01-JUL-1999: 99US-0141842.

PR 02-JUL-1999: 99US-0142154.

PR 06-JUL-1999: 99US-0142055.

PR 08-JUL-1999: 99US-0142390.

PR 09-JUL-1999: 99US-0142803.

PR 12-JUL-1999: 99US-0142977.

PR 13-JUL-1999: 99US-0143542.

PR 14-JUL-1999: 99US-0143624.

PR 15-JUL-1999: 99US-0144005.

PR 16-JUL-1999: 99US-0144085.

PR 16-JUL-1999: 99US-0144086.

PR 19-JUL-1999: 99US-0144325.

PR 19-JUL-1999: 99US-0144331.

PR 19-JUL-1999: 99US-0144332.

PR 19-JUL-1999: 99US-0144333.

PR 19-JUL-1999: 99US-0144334.

PR 20-JUL-1999: 99US-0144335.

PR 20-JUL-1999: 99US-0144632.

PR 20-JUL-1999: 99US-0144884.

PR 21-JUL-1999: 99US-0144814.

PR 21-JUL-1999: 99US-0145086.

PR 21-JUL-1999: 99US-0145088.

PR 22-JUL-1999: 99US-0145085.

PR 22-JUL-1999: 99US-0145087.

PR 22-JUL-1999: 99US-0145089.

PR 22-JUL-1999: 99US-0145192.

PR 23-JUL-1999: 99US-0145145.

PR 23-JUL-1999: 99US-0145218.

PR 23-JUL-1999: 99US-0145224.

PR 25-JUL-1999: 99US-0145276.

PR 27-JUL-1999: 99US-0145913.

PR 27-JUL-1999: 99US-0145918.

PR 27-JUL-1999: 99US-0145919.

PR 28-JUL-1999: 99US-0145951.

PR 02-AUG-1999: 99US-0146386.

PR 02-AUG-1999: 99US-0146388.

PR 02-AUG-1999: 99US-0146389.

PR 03-AUG-1999: 99US-0147038.

PR 04-AUG-1999: 99US-0147204.

PR 04-AUG-1999: 99US-0147302.

PR 05-AUG-1999: 99US-0147192.

PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148664.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
Query Match 86.8%; Score 33; DB 21; Length 257;

Best Local Similarity 100.0%; Pred. No. 64;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Caps 0;
QY 1 EKKLM 5
11111
Db 10 fKplw 14
RESULT 12
AAC15287
ID AAC15287 standard; Protein: 274 AA.
XX
AC AAC15287;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 15483.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
FE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 03-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151348.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 86.8%; Score 33; DB 21; Length 274;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1 FKPLM 5
DB 27 FKPLM 31

RESULT 13
AAC15286
ID AAC15286 standard; Protein: 279 AA.
XX
AC AAC15286;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 15482.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144315.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.

```

PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match      86.8%; Score 33; DB 21; Length 279;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 FKPLM 5
Db 32 Fkplm 36

```

```

RESULT 14
AAB01208
ID AAB01208 standard; Protein: 434 AA.
XX
AC AAB01208;
XX

```

```

DT 12-DEC-2000 (first entry)
XX
XX Corn putative lecithin:cholesterol acyltransferase #3.
DE
XX
XX Corn; lecithin:cholesterol acyltransferase; phytoosterol;
KW phosphatidylcholine-sterol O-acyltransferase; heat shock; cold shock.
XX
XX Zea mays.
OS
XX
XX WO200032791-A2.
PN
XX
XX 08-JUN-2000.
PD
XX
XX 02-DEC-1999; 99MO-US28586.
PF
XX
XX 03-DEC-1998; 98US-0110782.
HR
XX
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
PA
XX
XX Cahoon RE, Kinney AJ, Sakai H, Shen JB, Butler KH, Saylor JJ;
PI
XX
XX WPI: 2000-412337/35.
DR
XX
XX N-PSDB; AAA49203.
CR
XX
XX Polynucleotide encoding plant lecithin:cholesterol acyltransferase
PT enzyme useful for producing transgenic plants and for producing
PT antibodies specific to which is useful for screening cDNA expression
libraries
XX
XX
XX Claim 10; Page 40-41; 49pp; English.
PS
XX
XX The present sequence is a putative protein sequence of a corn
CC lecithin:cholesterol acyltransferase (also known as
CC phosphatidylcholine-sterol O-acyltransferase). This enzyme is found
CC associated with high-density lipoproteins and Apolipoprotein AI and -D.
CC The gene and protein can be used to produce transgenic plants which have
CC increased lipid metabolism and membrane fluidity, and therefore increased
CC resistance to heat and/or cold shock, to alter the content of phytoosterol
CC or lecithin in grains and to identify potential herbicides.
XX
XX
XX Sequence 434 AA:

```

```

Query Match      86.8%; Score 33; DB 21; Length 434;
Best Local Similarity 66.7%; Pred. No. 11e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 FKPLMK
Db 271 fkpmlk

```

```

RESULT 15
AAV35425
ID AAV35425 standard; Protein: 629 AA.
XX
AC AAV35425;
XX

```

```

DT 13-SEP-1999 (first entry)
XX
XX Protein involved in transcription, translation and/or maturation.
DE
XX
XX Respiratory disease: pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KM vaccine; neutralising epitope.
XX
XX Chlamydia pneumoniae.
OS
XX
XX WO9927105-A2.
PN
XX
XX 03-JUN-1999.
PD
XX
XX 20-NOV-1998; 98WO-IB01890.
PF

```

XX 04-NOV-1998: 98US-0107078.
 PR 21-NOV-1997: 97FR-0014673.

XX (GEST) GENSET.

XX Grifffais R;

XX WPI: 1999-357842/30.

XX Genome sequence of Chlamydia pneumoniae

XX Page 1207-1309; Disclosure: 1912pp; English.

XX AAY34584-y35879 represent the proteins encoded by all the open reading
 CC frames in the complete genome (see AAX9190) of Chlamydia pneumoniae.
 CC C. pneumoniae causes respiratory disease such as pneumonia and
 CC bronchitis and is thought to be a contributing factor in heart
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading
 CC frames of the C. pneumoniae genome (see AAY34584-y35879) can be used in
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
 CC nucleotide sequences can also be used as immunogenic compositions,
 CC especially where the vector directs the expression of a neutralising
 CC epitope of C. pneumoniae.

XX SQ Sequence 629 AA;

Query Match 84.2%; Score 32; DB 20; Length 629;

Best Local Similarity 83.3%; Pred. No. 2.5e+02;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FKPLWR 6
 I I I I I
 Db 290 fapIwr 295

Search completed: February 27, 2002, 11:41:16
 Job time: 455 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:45:19 : Search time 145.23 seconds
(without alignments)
3.147 Million cell updates/sec

Title: US-09-446-109A-10
Perfect score: 38
Sequence: 1 FKPLMR 6

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

1: PIR_68:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	94.7	623	2	peptide transport
2	35	92.1	380	2	formate hydrogently
3	35	92.1	462	2	hypothetical 47.5
4	35	92.1	462	2	probable amino act
5	35	92.1	2204	1	genome polyprotein
6	34	89.5	217	2	yedg protein - Esc
7	34	89.5	222	2	hypothetical prote
8	33	86.8	162	2	hypothetical prote
9	33	86.8	174	2	hypothetical prote
10	33	86.8	227	2	hypothetical prote
11	33	86.8	273	2	hypothetical prote
12	33	86.8	896	2	probable transloca
13	32	84.2	119	2	hypothetical prote
14	32	84.2	200	2	ribosomal protein
15	32	84.2	259	2	moda protein - pha
16	32	84.2	301	2	hypothetical prote
17	32	84.2	377	2	hypothetical prote
18	32	84.2	416	2	probable alkane hy
19	32	84.2	470	1	hypothetical prote
20	32	84.2	470	1	hypothetical 53k p
21	32	84.2	498	2	probable regulator
22	32	84.2	605	2	glucose transport
23	32	84.2	635	2	glucose transport
24	32	84.2	635	2	regulatory protein
25	32	84.2	664	2	Na+/glucose cotran
26	32	84.2	665	2	Na+/glucose cotran
27	32	84.2	700	2	DNA ligase - Deino
28	32	84.2	956	2	hypothetical prote
29	32	84.2	1048	2	Ca2+-transporting

30	32	84.2	1224	2	gene cob intron 3
31	32	84.2	2825	2	Dock protein, stre
32	31	81.6	233	1	envelope protein E
33	31	81.6	267	2	hypothetical prote
34	31	81.6	350	1	complement C5a ana
35	31	81.6	351	1	complement C5a ana
36	31	81.6	378	1	probable glycosyl
37	31	81.6	482	2	probable membrane
38	31	81.6	482	2	G protein-coupled
39	31	81.6	488	2	probable glucose-6
40	31	81.6	1007	2	hypothetical prote
41	30	78.9	222	2	hypothetical prote
42	30	78.9	242	2	conserved hypothet
43	30	78.9	259	1	corticotropin / 11
44	30	78.9	272	2	hypothetical prote
45	30	78.9	297	2	hypothetical prote

ALIGNMENTS

RESULT 1
T18250
peptide transport protein - yeast (Candida albicans)
C:Species: Candida albicans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18250
E:Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, November 1998
A:Reference number: Y18831
A:Accession: T18250
A:Status: preliminary: translated from CB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-623 <BAR>
A:Cross-references: EMBL:AL033503; NID:e1341066; PID:e1341077; PIDN:CAA22021.1
C:Genetics:
A:Gene: ptr2
A:Introns: 475/3

Query Match 94.7% Score 36; DB 2; Length 623;
Best Local Similarity 83.3% Pred. No. 16;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Cy 1 FKPLMR 6
DB 445 FKPLMR 450
RESULT 2
C54364
formate hydrogently, subunit 5 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: C54364
R:Butt, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sulten, G.G.; Blak
son, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: AF4300; MUD:96337999
A:Accession: C54364
A:Status: preliminary: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-380 <BUL>
A:Cross-references: CB:U67501; CB:L77117; NID:g2826289; PIDN:AA898504.1; PID:g1591218
C:Genetics:
A:Map position: REV458767-457625
C:Superfamily: NADH dehydrogenase (ubiquinone) 49k protein
Query Match 92.1% Score 35; DB 2; Length 380;

Best Local Similarity 83.3%; Pred. No. 15;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FKPLMR 6
DB 241 FKPLMR 246

RESULT 3

E65131
hypothetical 47.5 kD protein in cysG-trps intergenic region - Escherichia coli (strain K12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 17-Mar-2000
C:Accession: E65131
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shao, Y.
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: E65131
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1462 <BLAT>
A:Cross-references: GB:AE00413; GB:U00096; NID:g2367215; PIDN:AAC76395.1; PID:g2367216;
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: yhfM
C:Superfamily: arginine permease

Query Match 92.1%; Score 35; DB 2; Length 462;
Best Local Similarity 83.3%; Pred. No. 19;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 FKPLMR 6
DB 398 YKPLMR 403

RESULT 4

B86002
probable amino acid/amine transport protein yhfM [imported] - Escherichia coli (strain K12)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: B86002
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoustis, K.; Apodaca, Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B86002
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1462 <STO>
A:Cross-references: GB:AE005174; NID:g12517998; PIDN:AAG58478.1; GSPDB:GN00145; UMGF:247
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: yhfM

Query Match 92.1%; Score 35; DB 2; Length 462;
Best Local Similarity 83.3%; Pred. No. 19;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 FKPLMR 6
DB 398 YKPLMR 403

RESULT 5

RRNZNV
genome polypeptide - Newcastle disease virus (strain Beaudette C)
N:Contains: RNA-directed RNA polymerase (EC 2.7.7.48)

C:Species: Newcastle disease virus
C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 11-Jun-1999
C:Accession: A26747
R:Yusoff, K.; Millar, N.S.; Chambers, P.; Emmerson, P.T.

Nucleic Acids Res. 15, 3961-3976, 1987
A:Title: Nucleotide sequence analysis of the L gene of Newcastle disease virus: homology with the L gene of influenza A virus RNA polymerase
A:Reference number: A93655; MUID:87230982
A:Accession: A26747
A:Molecule type: mRNA
A:Residues: 1-2204 <YUS>
A:Cross-references: GB:A05399; NID:g60937; PIDN:CAA28985.1; PID:g60939
C:Genetics:
A:Gene: L
C:Superfamily: parainfluenza virus RNA-directed RNA polymerase
C:Keywords: ATP; nucleotidyltransferase

Query Match 92.1%; Score 35; DB 1; Length 2204;
Best Local Similarity 83.3%; Pred. No. 89;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FKPLMR 6
DB 1844 FKPLMR 1849

RESULT 6

H64956
ydcG protein - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 28-Jul-2000
C:Accession: H64956
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shao, Y.
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: H64956
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-217 <BLAT>
A:Cross-references: GB:AE00285; GB:U00096; NID:g1788229; PIDN:AAC74998.1; PID:g17882
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: ydcG
C:Superfamily: Escherichia coli conserved hypothetical protein ydcG

Query Match 89.5%; Score 34; DB 2; Length 217;
Best Local Similarity 83.3%; Pred. No. 14;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FKPLMR 6
DB 87 FKPLMR 92

RESULT 7

F85810
hypothetical protein yedG [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: F85810
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoustis, K.; Apodaca, Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: F85810
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-222 <STO>
A:Cross-references: GB:AE005174; NID:g12516032; PIDN:AAG56946.1; GSPDB:GN00145; UMGF:

A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: YedK
 C:Superfamily: Escherichia coli conserved hypothetical protein yedK

Query Match 89.5%; Score 34; DB 2; Length 222;
 Best Local Similarity 83.3%; Pred. No. 14;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FKPLM 6
 |||||
 DB 87 FKPLM 92

RESULT 8

S28301
 Hypothetical protein C40H1.6 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 26-Aug-1999
 C:Accession: S28301
 R:Berks, M.
 Submitted to the EMBL Data Library, December 1992
 A:Reference number: S28296
 A:Accession: S28301
 A:Molecule type: DNA
 A:Residues: 1-162 <BER>
 A:Cross-references: EMBL:Z19154; NID:g6650; PID:g6656
 C:Genetics:
 A:Introns: 32/3; 54/2; 107/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein C40H1.6

Query Match 86.8%; Score 33; DB 2; Length 162;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FKPLM 5
 |||||
 DB 120 FKPLM 124

RESULT 9

D86400
 Hypothetical protein AAD4599.1 [Imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: D86400
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpspiel, N.A.; Kaul, S.; White, O.; Alonso, Chlu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Jensen, N.F.; Hughes, B.; Hulzar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucos, J.S.; Malt, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: D86400
 A:Molecule type: DNA
 A:Status: preliminary
 A:Residues: 1-174 <STO>
 A:Cross-references: GB:AE005172; NID:g5668764; PIDN:AAD4599.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1
 C:Superfamily: Caenorhabditis elegans hypothetical protein C40H1.6

Query Match 86.8%; Score 33; DB 2; Length 174;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FKPLM 3
 |||||
 DB 124 FKPLM 128

RESULT 10

B70438
 Hypothetical protein aq_1596 - Aquifex aeolicus
 C:Species: Aquifex aeolicus
 C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
 C:Accession: B70438
 R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; V.
 Nature 392, 353-358, 1998
 A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 A:Reference number: A70300; MUID:98196666
 A:Accession: B70438
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-227 <ADP>
 A:Cross-references: GB:AE000747; NID:g2983944; PIDN:AC07503.1; PID:g2983955; GB:AE00
 A:Experimental source: strain VFS
 C:Genetics:
 A:Gene: aq_1596

Query Match 86.8%; Score 33; DB 2; Length 227;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FKPLM 5
 |||||
 DB 6 FKPLM 10

RESULT 11

T37841
 Probable translocation protein - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: T37841
 R:Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
 submitted to the EMBL Data Library, September 1997
 A:Reference number: 221749
 A:Accession: T37841
 A:Status: preliminary; translated from GB/EMBL/DDBU
 A:Molecule type: DNA
 A:Residues: 1-273 <MUR>
 A:Cross-references: EMBL:Z99162; PIDN:CA816220.1; GSPDB:GN00066; SPDB:SPAC17G6.09
 A:Experimental source: strain 972h-; cosmid C17G6
 C:Genetics:
 A:Gene: SPDB:SPAC17G6.09
 A:Map position: 1

Query Match 86.8%; Score 33; DB 2; Length 273;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FKPLM 5
 |||||
 DB 213 FKPLM 217

RESULT 12

F96523
 Hypothetical protein FLIA17.9 [Imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: F96523
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpspiel, N.A.; Kaul, S.; White, O.; Alonso, Chlu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, Jensen, N.F.; Hughes, B.; Hulzar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marzalli,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A6141; MUID:21016719
A:Accession: F96523
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-896 <STO>
A:Cross-references: GB:AE005173; NID:98778959; PID:AA049761.2; GSPDB:GN00141
C:Genetics:
A:Gene: F11A17.9
A:Map position: 1

Query Match 86.8%; Score 33; DB 2; Length 896;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KPLW 5
IIII
DB 255 KPLW 259

RESULT 13

S75418
ribosomal protein L29 - Sulfolobus solfataricus
N:Alternate names: protein C05001
C:Species: Sulfolobus solfataricus
C:Date: 11-Oct-1997 #sequence_revision 24-Oct-1997 #text_change 20-Jun-2000
C:Accession: S75418
R:Seisen, C.W.; Klenk, H.P.; Singh, R.K.; Allard, G.; Chan, C.C.Y.; Liu, Q.Y.; Penny, S.
Mol. Microbiol. 22, 175-191, 1996
A:Title: Organizational characteristics and information content of an archaeal genome: 1
A:Reference number: S73076; MUID:97055432
A:Accession: S75418
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-119 <SEN>
A:Cross-references: EMBL:Y08257; NID:91707772; PID:CAA69532.1; PID:91707824
A:Experimental source: strain P2
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1996
C:Superfamily: Haloarcula ribosomal protein HL29
C:Keywords: protein biosynthesis; ribosome

Query Match 84.2%; Score 32; DB 2; Length 119;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KPLW 6
IIII
DB 23 KPLW 27

RESULT 14

T10143
moda protein - phage T4
N:Alternate names: gp moda
C:Species: phage T4
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 23-Jul-1999
C:Accession: T10143; J20006
R:Frazier, M.W.; Mostig, G.
Gene 88, 7-14, 1990
A:Title: The bacteriophage T4 gene mri whose product inhibits late T4 gene expression in
A:Reference number: JH0136; MUID:90255970
A:Accession: T10143
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-200 <FRA>

A:Cross-references: EMBL:M30001; NID:91532074; PID:91532085

A:Experimental source: strain GT7
R:Mostig, G.; Colowick, N.C.
Personal communication, 1993
A:Reference number: J20005
A:Accession: J20006
A:Molecule type: DNA
A:Residues: 1-200 <MOS>
A:Note: submitted to T4 project
C:Genetics:
A:Gene: moda
A:Map position: 11.914-12.514
C:Function:
A:Description: probably involved in adenosylribosylation

Query Match 84.2%; Score 32; DB 2; Length 200;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KPLW 6
IIII
DB 68 KPLW 72

RESULT 15

E84293
hypothetical protein Vng1388h [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84293
R:Ng, W.V.; Kennedy, S.P.; Maharas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.;
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483
A:Accession: E84293
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-259 <STO>
A:Cross-references: GB:AE004437; NID:910580895; PID:AA019713.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNG1388H

Query Match 84.2%; Score 32; DB 2; Length 259;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KPLW 6
IIII
DB 95 KPLW 99

Search completed: February 27, 2002, 11:45:20
Job time: 699 sec

QY 1 FRPLMR 6
 ID 111:11
 DB 445 FRPLMR 450

RESULT 2
 YHFM_ECOLI STANDARD; PRT: 445 AA.
 AC P45359: P76686: 1-NOV-1995 (Rel. 32, Created)
 DT 01-JUL-1998 (Rel. 36, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOHETICAL 47.5 KDA PROTEIN IN CYSG-TRPS INTERGENIC REGION.
 GN YHFM OR B3370.
 OS Escherichia coli.
 OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE-97426617: PubMed-9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12."
 RL Science 277:1453-1474(1997).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
 CC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE ARCD/CADB/GERAB/LYXI FAMILY OF
 CC PERMEASIS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: U18997: AAA58167.1; ALT_SEQ.
 CC EMBL: AE000413: AAC76395.1; ALT_INIT.
 CC Ecogene: EC12908: yhfM.
 DR InterPro: IPR002293: AA_rel_permease_1.
 DR InterPro: IPR002027: Amino_acid_permease.
 DR Pfam: PF00324: aa_permeases; 1.
 KW Hypothetical protein; Transport; Transmembrane; Inner membrane;
 KW Complete proteome.
 FT TRANSMEM 10 30 POTENTIAL.
 FT TRANSMEM 38 58 POTENTIAL.
 FT TRANSMEM 93 113 POTENTIAL.
 FT TRANSMEM 121 141 POTENTIAL.
 FT TRANSMEM 155 175 POTENTIAL.
 FT TRANSMEM 181 201 POTENTIAL.
 FT TRANSMEM 236 256 POTENTIAL.
 FT TRANSMEM 273 293 POTENTIAL.
 FT TRANSMEM 334 354 POTENTIAL.
 FT TRANSMEM 355 375 POTENTIAL.
 FT TRANSMEM 389 410 POTENTIAL.
 FT TRANSMEM 417 435 POTENTIAL.
 SQ SEQUENCE 445 AA: 47576 MW: 9980594.45A5E1EEF CRC64;

Query Match 92.1%; Score 35; DB 1; Length 445;
 Best Local Similarity 83.3%; Pred. No. 11;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FRPLMR 6
 ID 111:11
 DB 381 YKPLMR 386

RESULT 3
 RRPL_NDV8 STANDARD; PRT: 2204 AA.
 AC P11205:
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE RNA POLYMERASE BETA SUBUNIT (EC 2.7.7.48) (LARGE STRUCTURAL PROTEIN)
 DE (L PROTEIN).
 GN L.
 OS Newcastle disease virus (strain Beaudette C/45) (NDV).
 OC Viruses: ssRNA negative-strand viruses; Mononegavirales;
 OC Paramyxoviridae; Paramyxovirinae; Rubulavirus.
 OX NCBI_TaxID=11178;
 RP SEQUENCE FROM N.A.
 RC MEDLINE-87230982: PubMed-3035486;
 RA Yusoff K., Millar N.S., Chambers P., Emerson P.T.;
 RT "Nucleotide sequence analysis of the L gene of Newcastle disease
 RT virus: homologues with Sendai and vesicular stomatitis viruses";
 RL Nucleic Acids Res. 15:3961-3976(1987).
 CC -1- FUNCTION: PROBABLE COMPONENT OF THE ACTIVE POLYMERASE. IT MAY
 CC SYNTHESIZE IN RNA SYNTHESIS, CAPING, METHYLATION AND POLY(A)
 CC GENE TRANSCRIPT, AND PROTEIN KINASE ACTIVITY.
 CC -1- SIMILARITY: WITH L PROTEIN OF OTHER PARAMYXOVIRUSES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X05399: CAA28985.1;
 CC PIR: A26747; RNNZNV.
 DR InterPro: IPR001016: Paramyx_RNA_POL.
 DR Pfam: PF00946: Paramyx_RNA_POL; 1.
 KW Transferrase; RNA-directed RNA polymerase.
 SQ SEQUENCE 2204 AA: 248822 MW: C678867AD904802C CRC64;

Query Match 92.1%; Score 35; DB 1; Length 2204;
 Best Local Similarity 83.3%; Pred. No. 49;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FRPLMR 6
 ID 111:11
 DB 1844 FRPLMR 1049

RESULT 4
 YEDK_ECOLI STANDARD; PRT: 217 AA.
 AC P76318:
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOHETICAL 24.5 KDA PROTEIN IN AMYA-FILE INTERGENIC REGION.
 GN YEDK OR B1931.
 OS Escherichia coli.
 OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RP SEQUENCE FROM N.A.
 RC STRAIN-JA11;
 RX MEDLINE-93381452: PubMed-8371104;
 RA Raha M., Kihara M., Kawagishi I., Macnab R.M.;
 RT "Organization of the Escherichia coli and Salmonella typhimurium
 RT chromosomes between flagellar regions IIIa and IIId, including a


```

RT large non-coding region."
RL J. Gen. Microbiol. 139:1401-1407(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655.
RX MEDLINE=9742617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.:
RT "The complete genome sequence of Escherichia coli K-12."
RL Science.277:1453-1474(1997).
CC -1- SIMILARITY: TO YEAST YMR114C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL: L13279; NOT_ANNOTATED_CDS.
DR EMBL; AE000285; AAC74998.1; -.
DR FcoGene; EG13278; yedk.
DR InterPro; IPR003738; DUF159.
DR Pfam; PF02586; DUF159; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 213 217 TRANS -> NC04ELI0PV (IN REF. 1).
SQ SEQUENCE 217 AA; 24500 MW; 4F26C95DB3B02900 CRC64;

OY 1 FKPLMR 6
    11111
Db 87 FKPLMQ 92

RESULT 5
HSPD_BRAJA STANDARD; PRT; 151 AA.
AC 069241;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE SMALL HEAT SHOCK PROTEIN HSPD.
GN HSPD.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bradyrhizobium group; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RA Narberhaus F., Weiglhofer W., Fischer H.M., Hennecke H.:
RT "Identification of the Bradyrhizobium japonicum deap gene as part of
RT an operon containing small heat shock protein genes."
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL: AJ003064; CAA05835.1; -.
DR InterPro; IPR002068; Crystallin_HSP20.

```

```

DR Pfam; PF00011; HSP20; 1.
DR PROSITE; PS01031; HSP20; 1.
KW Heat shock; Multigene family.
SQ SEQUENCE 151 AA; 17272 MW; FB44EF94FB599EE4 CRC64;

OY 1 FKPLMR 6
    11111
Db 6 FSPLMR 11

Query Match
Best Local Similarity 86.8%; Score 33; DB 1; Length 151;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPLMR 6
    11111
Db 6 FSPLMR 11

RESULT 6
HSPH_BRAJA STANDARD; PRT; 151 AA.
AC 086110;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE SMALL HEAT SHOCK PROTEIN HSPH.
GN HSPH.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bradyrhizobium group; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RA Muenchbach M., Nocker A., Narberhaus F.:
RT "Occurrence of a superfamily of small heat shock proteins in
RT Bradyrhizobium japonicum and other Rhizobium species: a plant-like
RT phenomenon."
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL: AJ010144; CAA09014.1; -.
DR InterPro; IPR002068; Crystallin_HSP20.
DR Pfam; PF00011; HSP20; 1.
DR PROSITE; PS01031; HSP20; 1.
DE Heat shock; Multigene family.
SQ SEQUENCE 151 AA; 17098 MW; DFE9809310AF6A28 CRC64;

OY 1 FKPLMR 6
    11111
Db 6 FSPLMR 11

Query Match
Best Local Similarity 86.8%; Score 33; DB 1; Length 151;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPLMR 6
    11111
Db 6 FSPLMR 11

RESULT 7
YLF6_CAEEL STANDARD; PRT; 162 AA.
AC 003598;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHEICAL 18.5 KDA PROTEIN C40H1.6 IN CHROMOSOME III.
GN C40H1.6.
OS Caenorhabditis elegans.

```

```

CC Eukaryota:Metazoa: Nematoda: Chromadorea: Rhabditida: Rhabditioidea:
OC Rhabditidae: Pelodolerinae: Caenorhabditis.
OX NCBI_TaxID=6239;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnscough R., Anderson K., Baynes C., Beets M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawling T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sultson J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Woldman P.;
RT *2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.*;
RL Nature.368:32-38(1994).
CC -1- SIMILARITY: STRONG, TO HUMAN CGI-126.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: 219154; CAA79557.1;
DR PIR: S28301; S28301.
DR WormPep: C40H1.6; CE00114.
KW Hypothetical protein.
SO SEQUENCE 162 AA: 18537 MW: 80C03CE3BC6E79D55 CRC64;

Query Match 86.8%; Score 33; DB 1; Length 162;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

OY 1 FRPLW 5
DB 120 FRPLW 124

RESULT 8
CCCG6_HUMAN STANDARD: PRT; 167 AA.
ID 1
AC O9Y3C8;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHEICAL PROTEIN CGI-126.
OS Homo sapiens (Human).
OC Eukaryota: Metazoa: Chordata: Vertebrata: Euteleostomi:
OC Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=20272150; PubMed=10810093;
RA Lai C.-H., Chou C.-Y., Ch'ang L.-Y., Liu C.-S., Lin W.-C.;
RT Identification of novel human genes evolutionarily conserved in
RT Caenorhabditis elegans by comparative proteomics.*;
RL Genome Res. 10:703-713(2000).
CC -1- SIMILARITY: STRONG, TO C.ELEGANS C40H1.6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: AF151884; AAD34121.1; -
CC KW Hypothetical protein.
CC SQ SEQUENCE 167 AA: 19458 MW: 167509187DCA3E14 CRC64:

      Query Match      86.8%  Score 33:  DB 1:  Length 167:
      Best Local Similarity 100.0%:  Pred. No. 11:
      Matches 5:  Conservative 0:  Mismatches 0:  Indels 0:  Gaps 0:

OY      1 FRKLW 5
          |||||
Nb      121 FRKLW 125

RESULT 9
R18E_SULSO
ID      R18E_SULSO      STANDARD:      PRT:      119 AA.
AC      P95990:
UT      01-NOV-1997 (Rel. 35, Created)
DT      20-AUG-2001 (Rel. 40, Last sequence update)
DE      50S RIBOSOMAL PROTEIN L18E.
GN      RPL18E OR SSC00070 OR C05001 OR C05_040.
OS      Sulfolobus solfataricus.
OC      Archaea: Crenarchaeota: Sulfolobales: Sulfolobaceae: Sulfolobus.
OX      NCBI_TaxID=2287;
RN      11)
RP      SEQUENCE FROM N.A.
RC      STRAIN=ATCC 35092 / DSM 1617 / P2:
RX      MEDLINE=9705432; PubMed=8899719;
RA      Sengen C.W., Kleink H.-P., Singh R.K., Allard G., Chan C.C.-Y.,
RA      Liu Q.Y., Penny S.L., Young F., Schenk M.E., Gaasterland T.,
RA      Doolittle W.F., Ragan M.A., Charlebois R.L.:
RA      "Organizational characteristics and information content of an
RA      archaeal genome: 156 kb of sequence from Sulfolobus solfataricus
RA      P2."
RA      Mol. Microbiol. 22:175-191(1996).
RL      [2]
RM      SEQUENCE FROM N.A.
RP      STRAIN=ATCC 35092 / DSM 1617 / P2:
RC      MEDLINE=21332296; PubMed=11427726;
RX      She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA      Aweyer M.J., Chan-Melher C.C.-Y., Clausen I.G., Curtis B.A.,
RA      De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA      Helkamp-de Jong T., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA      Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA      Charlebois R.L., Doolittle W.F., Duquet M., Gaasterland T.,
RA      Garrett R.A., Ragan M.A., Sengen C.W., Van der Oost J.:
RA      "The complete genome of the crenarchaeon Sulfolobus solfataricus P2."
RA      Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
RC      -I- SIMILARITY: BELONGS TO THE L18E FAMILY OF RIBOSOMAL PROTEINS.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      EMBL: Y08257; CAA69532.1; -
CC      DR EMBL: AL512962; CAC22781.1; -
CC      DR EMBL: AE006647; AAK40432.1; -
CC      DR InterPro: IPR000039; Ribosomal_L18e.
CC      DR Pfam: PF00256; L15; 1.
CC      DR PROSITE: PS01106; RIBOSOMAL_L18E; FALSE_NEG.
CC      KW Ribosomal protein, complete proteome.
CC      SQ SEQUENCE 119 AA: 13492 MW: 0F99A26B198AE2E CRC64:

```


DR SMART: SM00345; HTH_CNTR: 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 470 AA; 53046 MW; 0C65DA4DC69A4FDA CRC64;

Query Match 84.2%; Score 32; DB 1; Length 470;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KPLMR 6
 11111
 Db 122 KPLMR 126

RESULT 13
 ID SL51_PIC STANDARD; PRT: 605 AA.
 AC P26429;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE SODIUM/GLUCOSE COTRANSPORTER 1 (NA(+)/GLUCOSE COTRANSPORTER 1)
 DE (HIGH AFFINITY SODIUM-GLUCOSE COTRANSPORTER) (FRAGMENT).
 CN SLC5A1 OR SGLT1.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91061757; PubMed=2247068;
 RA Ohta T., Isselbacher K.J., Rhoads D.B.;
 RT "Regulation of glucose transporters in LLC-PK1 cells: effects of D-
 glucose and monosaccharides.";
 RL Mol. Cell. Biol. 10:6491-6499(1990).
 CC -1- FUNCTION: ACTIVELY TRANSPORTS GLUCOSE INTO CELLS BY NA+
 CO-TRANSPORT WITH A NA+ TO GLUCOSE COUPLING RATIO OF 2:1.
 CC -1- FUNCTION: EFFICIENT SUBSTRATE TRANSPORT IN MAMMALIAN KIDNEY IS
 PROVIDED BY THE CONCENTRATED ACTION OF A LOW AFFINITY HIGH CAPACITY
 AND A HIGH AFFINITY LOW CAPACITY NA(+)/GLUCOSE COTRANSPORTER
 ARRANGED IN SERIES ALONG KIDNEY PROXIMAL TUBULES.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: M34044; AAA31122.1;
 DR PIR: A36361; A36361.
 DR InterPro: IPR001734; Na_solut_symport.
 DR Pfam: PF00474; SSF: 1
 DR PROSITE: PS00456; NA_SOLUT_SYM_1; 1.
 DR PROSITE: PS00457; NA_SOLUT_SYM_2; 1.
 DR PROSITE: PS50283; NA_SOLUT_SYM_3; 1.
 KW Transport; Sugar transport; Transmembrane; Sodium transport; Symport;
 KW Glycoprotein.
 FT NON_TER 1 1
 FT DOMAIN <1 7 EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 8 28 POTENTIAL.
 FT DOMAIN 29 48 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 49 69 POTENTIAL.
 FT DOMAIN 70 114 EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 115 134 POTENTIAL.
 FT DOMAIN 135 151 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 152 172 POTENTIAL.
 FT DOMAIN 173 213 EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 214 234 POTENTIAL.
 FT DOMAIN 235 257 CYTOPLASMIC (POTENTIAL).

FT TRANSSEM 258 277 POTENTIAL.
 FT DOMAIN 278 366 EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 367 386 POTENTIAL.
 FT DOMAIN 387 398 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 399 419 POTENTIAL.
 FT DOMAIN 420 469 EXTRACELLULAR (POTENTIAL).
 FT TRANSSEM 470 490 POTENTIAL.
 FT DOMAIN 491 583 POTENTIAL.
 FT TRANSSEM 584 604 POTENTIAL.
 FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
 FT SITE 243 243 IMPLICATED IN SODIUM COUPLING
 (BY SIMILARITY).
 SQ SEQUENCE 605 AA; 66917 MW; 132B6D08E0199F02 CRC64;

Query Match 84.2%; Score 32; DB 1; Length 605;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KPLMR 6
 11111
 Db 579 KPLMR 583

RESULT 14
 ID SL51_HUMAN STANDARD; PRT: 664 AA.
 AC P13866;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE SODIUM/GLUCOSE COTRANSPORTER 1 (NA(+)/GLUCOSE COTRANSPORTER 1)
 DE (HIGH AFFINITY SODIUM-GLUCOSE COTRANSPORTER).
 CN SLC5A1 OR SGLT1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89145544; PubMed=2490366;
 RA Hediger M.A., Turk E., Wright E.M.;
 RT "Homology of the human intestinal Na+/glucose and Escherichia coli
 Na+/proline cotransporters.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:5748-5752(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Swan M.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20057165; PubMed=10591308;
 RA Dunham I., Hunt A.R., Collins J.E., Bruskiewicz R., Beare D.M.,
 RA Clamp M., Smink L.J., Alnsough R., Almeida J.P., Babage A.,
 RA Baguley C., Bailey J., Barlow K., Bates K.N., Beasley O., Bird C.P.,
 RA Blakey S., Bridgeman A.M., Buck D., Burgess J., Burrill W.D.,
 RA Burton J., Carder C., Carter N.P., Chen Y., Clark G., Clegg S.M.,
 RA Cobley V., Cole G.C., Collier R.E., Connor R.E., Conroy D., Corby N.,
 RA Coville G.J., Cox A.V., Davis J., Dawson E., Dhami P.D., Dockree C.,
 RA Dodsworth S.J., Durbin R.M., Ellington A., Evans K.L., Fey J.M.,
 RA Fleming K., French L., Garner A.A., Gilbert J.G.R., Goward M.E.,
 RA Gratnam D., Griffiths W.N., Hall C., Hall R., Hall-Tamlyn G.,
 RA Heathcote R.W., Ho S., Holmes S., Hunt S.E., Jones M.C., Kershaw J.,
 RA Kimberley A., King A., Laird G.K., Langford C.F., Leverish M.A.,
 RA Lloyd C., Lloyd D.M., Martyn I.D., Mashreghi-Mohammadi M.,
 RA Matthews L., McCann O.T., Mcclay J., McLaren S., Mcmurtry A.A.,
 RA Milne S.A., Mortimore B.J., Odell C.N., Pavitt R., Pearce A.V.,
 RA Pearson D., Phillips B.J., Phillips S.H., Plumb R.M., Ramsey H.,
 RA Ramsey Y., Rogers L., Ross M.T., Scott C.E., Sehra H.K., Skuce C.D.,
 RA Smalley S., Smith M.L., Soderlund C., Spragon L., Steward C.A.,
 RA Sulston J.E., Swan R.M., Vaudin M., Wall M., Wallis J.M.,
 RA Whiteley M.N., Willey D., Williams L., Williams S., Williamson H.,
 RA Wilmer T.E., Wilming L., Wright C.L., Hubbard T., Bentley D.R.,

RA Beck S., Rogers J., Shimizu N., Minoshima S., Kawasaki K., Sasaki T.,
RA Asakawa S., Kuoh J., Shintani A., Shibuya K., Yoshizaki Y., Aoki N.,
RA Mitsuyma S., Roe B.A., Chen F., Chu L., Crabtree J., Descamps S.,
RA Do A., Do T., Dorman A., Fang F., Fu Y., Hu P., Hua A., Kelson S.,
RA Lai H., Lao H.T., Lewis J., Lewis S., Lin S.-P., Loh P., Mala E.,
RA Nguyen T., Pan H., Phan S., Qi S., Qian Y., Ray L., Ren O., Shauli S.,
RA Sloan D., Song L., Wang Q., Wang Y., Wang Z., White J., Willingham D.,
RA Xu H., Yao Z., Zhan M., Zhang G., Chiswick S., Murray J., Miller N.,
RA Min P., Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H.,
RA Bourne J., Cordes M., Du Z., Fulton L., Goela D., Graves T.,
RA Hawkins J., Hinds K., Kemp K., Latreille P., Layman D., Ozerky P.,
RA Rohlfing T., Scheet P., Walker C., Wansley A., Wohlmann P., Pepin K.,
RA Nelson J., Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R.,
RA Nelson R., Emanuel B.S., Shaikh T., Kurahashi H., Salta S.,
RA Badger M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
RA Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
RA Peyrard M., Kedra D., Serousi E., Fransson I., Tapia I., Bruder C.E.,
RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
RA Khan A.S., Lane L., Tiliakun Y., Wright H.,
RT "The DNA sequence of human chromosome 22.";
RL Nature 402:489-495(1999).
RN [4]
RP VARIANT CGM ASN-28.
RX MEDLINE=91179516; Pubmed=2008213;
RA Turk E., Zabel B., Mundlos S., Dyer J., Wright E.M.;
RT "Glucose/galactose malabsorption caused by a defect in the
RT Na⁺/glucose cotransporter.";
RL Nature 350:354-356(1991).
CC -1- FUNCTION: ACTIVELY TRANSPORTS GLUCOSE INTO CELLS BY NA⁺.
CC -1- CO-TRANSPORT WITH A NA⁺ TO GLUCOSE COUPLING RATIO OF 2:1.
CC -1- FUNCTION: EFFICIENT SUBSTRATE TRANSPORT IN MAMMALIAN KIDNEY IS
CC PROVIDED BY THE CONCENTRATED ACTION OF A LOW AFFINITY HIGH CAPACITY
CC AND A HIGH AFFINITY LOW CAPACITY NA(+)/GLUCOSE COTRANSPORTER
CC ARRANGED IN SERIES ALONG KIDNEY PROXIMAL TUBULES.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- DISEASE: DEFECT IN THE CO-TRANSPORTER IS THE ORIGIN OF THE
CC CONGENITAL GLUCOSE-GALACTOSE MALABSORPTION SYNDROME (GCM).
CC -1- SIMILARITY: BELONGS TO THE SODIUM: Solute symporter family (SSF).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L293339; AAB59448.1; -;
DR EMBL: L293328; AAB59448.1; JOINED.
DR EMBL: L293330; AAB59448.1; JOINED.
DR EMBL: L293329; AAB59448.1; JOINED.
DR EMBL: L293331; AAB59448.1; JOINED.
DR EMBL: L293332; AAB59448.1; JOINED.
DR EMBL: L293333; AAB59448.1; JOINED.
DR EMBL: L293334; AAB59448.1; JOINED.
DR EMBL: L293335; AAB59448.1; JOINED.
DR EMBL: L293336; AAB59448.1; JOINED.
DR EMBL: L293337; AAB59448.1; JOINED.
DR EMBL: L293338; AAB59448.1; JOINED.
DR EMBL: M24847; AAB60320.1; -;
DR EMBL: AL022331; CA18445.2; -;
DR EMBL: Z83849; CAB06090.2; -;
DR EMBL: Z74021; CA98336.2; -;
DR EMBL: Z80998; CAB02632.2; -;
DR EMBL: Z83839; CAB06087.1; -;
DR EMBL: Z74021; CAB06087.1; JOINED.
DR EMBL: Z80998; CAB06087.1; JOINED.
DR EMBL: Z83849; CAB06087.1; JOINED.
DR EMBL: AL022331; CAB06087.1; JOINED.
DR PIR: A33545; A33545.
DR MIM: 182380; -;
DR InterPro: IPR001734; Na⁺-solut⁺-symport.
DR Pfam: PF00474; SSF. 1.

DR PROSITE: PS00456; NA_SOLUT_SYM_1; 1.
DR PROSITE: PS00457; NA_SOLUT_SYM_2; 1.
DR PROSITE: PS0283; NA_SOLUT_SYM_3; 1.
KW Transport: Sugar transport; Transmembrane; Sodium transport; Symport;
KW Glycoprotein; Disease mutation.
FT DOMAIN 1 28
FT PROSITE 29 47 POTENTIAL.
FT DOMAIN 48 64 EXTRACELLULAR (POTENTIAL).
FT PROSITE 65 85 POTENTIAL.
FT DOMAIN 86 105 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 106 126 POTENTIAL.
FT DOMAIN 127 171 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 172 191 POTENTIAL.
FT DOMAIN 193 208 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 209 229 POTENTIAL.
FT DOMAIN 230 270 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 271 291 POTENTIAL.
FT DOMAIN 292 314 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 315 334 POTENTIAL.
FT DOMAIN 335 423 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 424 443 POTENTIAL.
FT DOMAIN 444 455 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 456 476 POTENTIAL.
FT DOMAIN 477 526 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 527 547 POTENTIAL.
FT DOMAIN 548 642 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 643 663 POTENTIAL.
FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT VARIANT 28 28 D -> N (IN GCM).
FT SEQUENCE 664 AA: 73497 MW: 28403376595EAB74 CRC64;
FT 50
Query Match 84.2% Score 32; DR 1; Length 664;
Best Local Similarity 100.0% Pred. No. 60;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
(yy 2 KPLMR 6
11111
Db 638 KPLMR 642
RESULT 15
SL51_RAT STANDARD: PRT: 665 AA.
AC PS3790;
ID 01-OCT-1996 (Rel. 34, Created)
ID 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE SODIUM/GLUCOSE COTRANSPORTER 1 (NA(+)/GLUCOSE COTRANSPORTER 1)
DE (HIGH AFFINITY SODIUM-GLUCOSE COTRANSPORTER).
GN SLC5A1 OR SGLT1.
US Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Kidney;
RX MEDLINE=94216314; Pubmed=8163506;
RA Lee W.S., Kanai Y., Wells R.G., Hediger M.A.;
RT "The high affinity Na⁺/glucose cotransporter. Re-evaluation of
RT function and distribution of expression.";
RT J. Biol. Chem. 269:12032-12039(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Kasahara M., Mori K.;
RL Submitted (MAY-1993) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ACTIVELY TRANSPORTS GLUCOSE INTO CELLS BY NA⁺.
CC -1- CO-TRANSPORT WITH A NA⁺ TO GLUCOSE COUPLING RATIO OF 2:1.
CC -1- FUNCTION: EFFICIENT SUBSTRATE TRANSPORT IN MAMMALIAN KIDNEY IS
CC PROVIDED BY THE CONCENTRATED ACTION OF A LOW AFFINITY HIGH CAPACITY
CC AND A HIGH AFFINITY LOW CAPACITY NA(+)/GLUCOSE COTRANSPORTER

```
CC      ARRANGED IN SERIES ALONG KIDNEY PROXIMAL TUBULES.
CC      -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC      -1- DEVELOPMENTAL STAGE: APPEARS ON EMBRYONIC DAY 18 AND GRADUALLY
CC      INCREASES UNTIL BIRTH.
CC      -1- SIMILARITY: BELONGS TO THE SODIUM:SOLUTE SYMPORTER FAMILY (SSF).
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified; and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: U03120; AAA19015.1; -.
CC      EMBL: D16101; BAA03676.1; -.
CC      InterPro: IPR001734; Na_solut_symport.
CC      Pfam: PF00474; SSF: 1.
CC      PROSITE: PS00456; NA_SOLUT_SYMP_1; 1.
CC      PROSITE: PS00457; NA_SOLUT_SYMP_2; 1.
CC      PROSITE: PS0283; NA_SOLUT_SYMP_3; 1.
CC      TransPort: Sugar transport; Transmembrane; Sodium transport; Symport;
CC      Glycoprotein.
KW      DOMAIN 1 28      CYTOPLASMIC (POTENTIAL).
FT      TRANSMEM 29 47      POTENTIAL.
FT      DOMAIN 48 64      EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM 65 85      POTENTIAL.
FT      DOMAIN 86 105      CYTOPLASMIC (POTENTIAL).
FT      TRANSMEM 106 126      POTENTIAL.
FT      DOMAIN 127 171      EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM 172 191      POTENTIAL.
FT      DOMAIN 193 208      CYTOPLASMIC (POTENTIAL).
FT      TRANSMEM 209 229      POTENTIAL.
FT      DOMAIN 230 270      EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM 271 291      POTENTIAL.
FT      DOMAIN 292 314      CYTOPLASMIC (POTENTIAL).
FT      TRANSMEM 315 334      POTENTIAL.
FT      DOMAIN 335 423      EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM 424 443      POTENTIAL.
FT      DOMAIN 444 455      CYTOPLASMIC (POTENTIAL).
FT      TRANSMEM 456 476      POTENTIAL.
FT      DOMAIN 477 526      EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM 527 547      POTENTIAL.
FT      DOMAIN 548 643      CYTOPLASMIC (POTENTIAL).
FT      TRANSMEM 644 664      POTENTIAL.
FT      CARBOHYD 248 248      N-LINKED (GLCNAC: . .) (POTENTIAL).
SQ      SEQUENCE 665 AA; 73066 MW; A92038D964BFF061 CRC64;
```

Query Match 84.28; Score 32; DB 1; Length 665;
Best Local Similarity 100.08; Pred. No. 61;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 KPLMR 6
DB 639 KPLMR 643

Search completed: February 27, 2002, 11:42:46
Job time: 545 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:50:10 ; Search time 281.76 Seconds
(without alignments)
3.115 Million cell updates/sec

Title: US-09-446-109a-10
Sequence: 1 FKPLMR 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 14627329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP archaea: *
2: SP bacteria: *
3: SP fungi: *
4: SP human: *
5: SP_invertebrate: *
6: SP_mammal: *
7: SP_mhc: *
8: SP_organelle: *
9: SP_phage: *
10: SP_plant: *
11: SP_rodent: *
12: SP_virus: *
13: SP_vertebrate: *
14: SP_unclassified: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	100.0	917	5	Q9XPB8
2	38	100.0	917	5	Q9VY94
3	36	94.7	623	3	Q9URU7
4	35	92.1	380	1	057935
5	35	92.1	2204	12	090341
6	35	92.1	2204	12	09MMH6
7	35	92.1	2204	12	09DUD3
8	34	89.5	138	2	Q9ZEM4
9	34	89.5	262	5	09NND2
10	34	89.5	2093	5	09XZ8
11	33	86.8	164	5	09V7H5
12	33	86.8	167	4	09P009
13	33	86.8	167	4	09BS96
14	33	86.8	174	11	09CR09
15	33	86.8	174	10	09SXC8
16	33	86.8	175	11	09CT38
17	33	86.8	227	2	067532
18	33	86.8	250	2	09EUM2
19	33	86.8	273	3	013787

20	33	86.8	284	10	Q9FM24	Q9FM24 arabidopsis
21	33	86.8	361	4	Q15595	Q15595 homo sapien
22	33	86.8	361	4	Q9HAS5	Q9HAS5 homo sapien
23	33	86.8	416	4	Q9HAS6	Q9HAS6 homo sapien
24	33	86.8	416	4	Q15407	Q15407 homo sapien
25	33	86.8	896	10	Q9SXB9	Q9SXB9 arabidopsis
26	33	86.8	978	2	Q07831	Q07831 pseudomonas
27	33	86.8	978	2	Q9RA01	Q9RA01 pseudomonas
28	32	84.2	102	11	Q9CV08	Q9CV08 mus musculus
29	32	84.2	132	11	Q70449	Q70449 rattus norv
30	32	84.2	180	2	Q9KZM9	Q9KZM9 streptomyce
31	32	84.2	185	5	Q9N775	Q9N775 leishmania
32	32	84.2	232	4	Q75536	Q75536 homo sapien
33	32	84.2	238	11	Q9D847	Q9D847 mus musculus
34	32	84.2	254	11	Q9CR07	Q9CR07 mus musculus
35	32	84.2	259	1	Q9HC04	Q9HC04 halobacteri
36	32	84.2	301	5	Q22568	Q22568 caenorhabd
37	32	84.2	317	2	Q9LAE9	Q9LAE9 bacillus ce
38	32	84.2	325	2	Q9LAE9	Q9LAE9 bacillus ce
39	32	84.2	377	2	Q9I318	Q9I318 pseudomonas
40	32	84.2	416	2	Q05895	Q05895 mycobacteri
41	32	84.2	428	2	Q9RBM2	Q9RBM2 wolbachia s
42	32	84.2	438	2	Q83784	Q83784 treponema p
43	32	84.2	532	5	Q9VTS5	Q9VTS5 drosophila
44	32	84.2	635	2	Q9Z7B3	Q9Z7B3 chlamydia p
45	32	84.2	635	2	Q9JSA3	Q9JSA3 chlamydia p

ALIGNMENTS

RESULT 1
ID Q9XPB8 PRELIMINARY: PRT: 917 AA.
AC Q9XPB8;
BT 01-NOV-1999 (TEMBLrel. 12, Created)
DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)
HT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE GAMMA-TUBULIN RING PROTEIN OGRIP91.
CN L(1)DD4 OR DGRIP91 OR CG10988.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
CX Ephydroidea; Drosophilidae; Drosophila.
NC NCBL_TaxID=7227;
FN [1]
FM SEQUENCE FROM N.A.
FX MEDLINE=9156983; PubMed=10037793;
FA Oegema K., Wiese C., Martin O., Milligan R.A., Twamatsu A.,
KA Mitchison T.J., Zheng Y.;
KT "Characterization of two related Drosophila gamma-tubulin complexes
that differ in their ability to nucleate microtubules."
RL J. Cell Biol. 144:721-733(1999).
DR EMBL: AF118380; AAD27817.1;
DR FlyBase: FBgn0001612; l(1)dd4.
DR InterPro: IPR000634; dehydrtase_ser_thr.
DR PROSITE: PS00165; DEHYDRATASE_SER_THR; 1.
SQ SEQUENCE 917 AA: 103819 MW: 882241E35049D27 CRC64:

Query Match Best Local Similarity 100.0%; Score 38; DB 5; Length 917;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FKPLMR 6
bfb 657 FKPLMR 662
RESULT 2
Q9VY94 PRELIMINARY: PRT: 917 AA.
AC Q9VY94;

```

DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE GRP91 PROTEIN.
CN L(1)DD4 OR CG10988.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:
OC Pterygota: Neoptera: Endopterygota: Diptera: Brachycera: Muscomorpha:
OC Ephydroidea: Drosophilidae: Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Morten J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.C., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Agbayani A., An H.-J., Andrews-Plamkoc C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostlin D., Houston K.A., Howland T.J., Wei M.H., Idegawa C.,
RA Jallat M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matceli B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Pelazolo M., Plutman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003493; AAF48309.1;
DR FlyBase: FBgn0001612; 1(1)dd4.
DR InterPro: IPR000634; dehydratase ser. thr.
DR PROSITE: PS00165; DEHYDRATASE_SER_THR; 1.
SQ SEQUENCE 917 AA: 103706 MW: 6AEB86C211D256BB CRC64;

```

```

DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE PEPTIDE TRANSPORT PROTEIN.
CN PTR2.
OS Candida albicans (Yeast).
OC Eukaryota: Fungi: Ascomycota: Saccharomycotina: Saccharomycetes:
OC Saccharomycetales: mitosporic Saccharomycetales: Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1161;
RA Murphy L., Harris D.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1161;
RA Barrall B.G., Rajandream M.A.;
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=1161;
RX MEDLINE=97435544; PubMed=9290243;
RA Tait E., Simon M.C., King S., Brown A.J., Gow N.A.R., Shaw D.J.;
RT "A Candida albicans Genome Project: Cosmid Contigs, Physical Mapping,
RT and Gene Isolation."
RL Fungal Genet. Biol. 21:308-314(1997).
DR EMBL: AL033503; CA622021.1;
DR InterPro: IPR000109; PTR2.
DR Pfam: PF00854; PTR2; 1.
DR PROSITE: PS01022; PTR2-1; 1.
DR PROSITE: PS01023; PTR2-2; UNKNOWN.1.
SQ SEQUENCE 623 AA: 69943 MW: 8543A3B1F7E7363E CRC64;

```

```

Query Match          100.0%; Score 38; DB 5; Length 917;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FKPLMR 6
   11111
Db 657 FKPLMR 662

RESULT 3
Q9URJ7 PRELIMINARY; PRT: 623 AA.
AC Q9URJ7;
DT 01-MAY-2000 (TReMBLrel. 13, Created)

Query Match          94.7%; Score 36; DB 3; Length 623;
Best Local Similarity 83.3%; Pred. No. 63;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FKPLMR 6
   111:11
Db 445 FKPLMR 450

RESULT 4
Q57935 PRELIMINARY; PRT: 380 AA.
AC Q57935;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JUN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE PUTATIVE FORMATE HYDROXYLASE SUBUNIT 5 (FHL SUBUNIT 5) (HYDROGENASE-
DE 3 COMPONENT E).
OE M050515.
OS Methanococcus jannaschii.
OC Archaea: Euryarchaeota: Methanococcales: Methanococcaceae:
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.T.,
RA Overbeek R., Kirkness E.F., Weinstock K.C., Merrick J.M., Glodex A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fulmann J.L., Nguyen D.,
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Bordovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RL jannaschii."
RL Science 273:1058-1073(1996).
CC -i- COFACTOR: NICKEL (BY SIMILARITY).

```


CC -1- PATHWAY: HYDROGEN METABOLISM: FOL. PATHWAY (BY SIMILARITY).
 CC -1- SUBUNIT: FOL COMPRESES OF A FORMATE DEHYDROGENASE, UNIDENTIFIED
 CC ELECTRON CARRIERS AND A HYDROGENASE (ISOENZYME 3). IN THIS NON-
 CC ENERGY CONSERVING PATHWAY MOLECULAR HYDROGEN AND CARBOXYLATE FROM
 CC FORMATE ARE RELEASED (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE COMPLEX I 49 KDA SUBUNIT FAMILY.
 DR EMBL: U67501; AAB98504.1; -
 DR TIGR: M10515; -
 DR Interpro: IPR001135; ComplexI_49kd.
 DR Interpro: IPR001501; Nifese_Hases.
 DR Pfam: PF00346; complexI_49kd; 1.
 DR Pfam: PF00374; Nifese_Hases; 1.
 DR PROSITE: PS00535; COMPLEXI_49K; UNKNOWN_1.
 DR PROSITE: PS00507; N1_HCNASE.L1; UNKNOWN_1.
 KM Hypothetical protein: Oxidoreductase; NAD; Iron-sulfur; 4Fe-4S;
 KM Nickel; Complete proteome
 SQ SEQUENCE 380 AA; 43638 MW; 3EAD16D67834A76F CRC64;

Query Match 92.1%; Score 35; DB 1; Length 380;
 Best Local Similarity 83.3%; Pred. No. 61;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FKPLMR 6
 Db 241 FKPYMR 246

RESULT 5
 ID 090341 PRELIMINARY; PRT; 2204 AA.
 AC 090341;
 DT 01-NOV-1998 (Tremblrel. 08, Created)
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE LARGE POLYMERASE PROTEIN.
 GN L.
 OS Newcastle disease virus.
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
 OC Paramyxoviridae; Paramyxovirinae; Rubulavirus.
 OX NCBI_TaxID=11176;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LASOTA;
 RA de Leeuw O.S., Peeters B.P.H.;
 RT "Complete nucleotide sequence of Newcastle disease virus: evidence for
 RT the existence of a new genus within the subfamily Paramyxovirinae.";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF077761; AAC28375.1; -
 DR Interpro: IPR001016; Paramyx-RNA_pol.
 DR Pfam: PF00946; Paramyx-RNA_pol; 1.
 SQ SEQUENCE 2204 AA; 248675 MW; FCI7CD12F266E6E4 CRC64;

Query Match 92.1%; Score 35; DB 12; Length 2204;
 Best Local Similarity 83.3%; Pred. No. 3.3e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FKPLMR 6
 Db 1844 FRPLMR 1849

RESULT 6
 ID 09MMH6 PRELIMINARY; PRT; 2204 AA.
 AC 09MMH6;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE LARGE PROTEIN.
 GN L.
 OS Newcastle disease virus.

OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
 CC Paramyxoviridae; Paramyxovirinae; Rubulavirus.
 OX NCBI_TaxID=11176;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Roemer-Oberdorfer A., Buchholz U.J., Mundt E., Mettenleiter T.C.;
 RT "Generation of recombinant lentogenic Newcastle Disease Virus from
 RT cDNA.";
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Roemer-Oberdorfer A.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Y18898; CAB51327.1; -
 DR Interpro: IPR001016; Paramyx-RNA_pol.
 DR Pfam: PF00946; Paramyx-RNA_pol; 1.
 SQ SEQUENCE 2204 AA; 248500 MW; 5E1F506D0E608EDD CRC64;

Query Match 92.1%; Score 35; DB 12; Length 2204;
 Best Local Similarity 83.3%; Pred. No. 3.3e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FKPLMR 6
 Db 1844 FRPLMR 1849

RESULT 7
 ID 09DL03 PRELIMINARY; PRT; 2204 AA.
 AC 09DL03;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE LARGE POLYMERASE PROTEIN.
 GN L.
 OS Newcastle disease virus.
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
 OC Paramyxoviridae; Paramyxovirinae; Rubulavirus.
 OX NCBI_TaxID=11176;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BI;
 RA Sellars H.S., Seal B.S.;
 RT "Complete sequence for the BI strain of Newcastle disease virus.";
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF309418; AAC36980.1; -
 DR Interpro: IPR001016; Paramyx-RNA_pol.
 DR Pfam: PF00946; Paramyx-RNA_pol; 1.
 SQ SEQUENCE 2204 AA; 248738 MW; 9683688FD495A2D2 CRC64;

Query Match 92.1%; Score 35; DB 12; Length 2204;
 Best Local Similarity 83.3%; Pred. No. 3.3e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FKPLMR 6
 Db 1844 FRPLMR 1849

RESULT 8
 ID 09ZEM4 PRELIMINARY; PRT; 138 AA.
 AC 09ZEM4;
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE ORF1 PROTEIN.
 GN K1381211 pneumoniae.
 OS K1381211 pneumoniae.
 OC Plasmid pGS1500.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Klebsiella.
 OX NCBI_TaxID=573;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=APP;
 RA Osborn A.M., Da Silva Taley F.M., Steyn L.M., Pickup R.M.,
 SA Saunders J.R.;
 RT "Mosaic plasmids and mosaic replicons: evolutionary lessons from the
 RL analysis of genetic diversity in IncFII-related replicons";
 DR EMBL: A009980; CA08926.1;
 DR InterPro: IPR003738; DUF159.
 DR Pfam: PF02586; DUF159; 1.
 KW Plasmid.
 SQ SEQUENCE 138 AA: 15358 MW: E450AF354517EED9 CRC64;

Query Match 89.5%; Score 34; DB 2; Length 138;
 Best Local Similarity 83.3%; Pred. No. 35;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FKPLMR 6
 DB 2 FKPLMQ 7

RESULT 9
 O9NND2 PRELIMINARY: PRT: 262 AA.

AC O9NND2;
 DT 01-OCT-2000 (TEMBLrel. 15, Created)
 DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TEMBLrel. 15, Last annotation update)
 DE PROBABLE AMP DEAMINASE (FRAGMENT).
 GN LML3.261.
 OS Leishmania major.
 OC Eukaryota; Metazoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5664;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FRIDLIN;
 RA Murphy L., Quail M., Harris D., Rajandream M., Ivens A., Barrell B.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL160371; CAC00278.1;
 FT NON_TER 1 1
 FT NON_TER 262 262
 SQ SEQUENCE 262 AA: 30458 MW: ADE667C85EB700B CRC64;

Query Match 89.5%; Score 34; DB 5; Length 262;
 Best Local Similarity 83.3%; Pred. No. 65;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FKPLMR 6
 DB 82 FEPLMR 87

RESULT 10
 O9X2Y8 PRELIMINARY: PRT: 2093 AA.
 AC O9X2Y8;
 DT 01-NOV-1999 (TEMBLrel. 12, Created)
 DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
 DE AMP DEAMINASE.
 GN L302.11.
 OS Leishmania major.
 OC Eukaryota; Metazoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5664;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FRIDLIN;

KA Ivens A.C., Murphy L., Quail M., Harris D., Oliver K., Lawson D.,
 RA Rajandream M.A., Barrell B.C.;
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FRIDLIN;
 RX MEDLINE=98146435; PubMed=9477341;
 RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
 RA Smith D.F.;
 RT "A physical map of the Leishmania major Friedlin genome";
 RL Genome Res. 8:133-145(1998).
 DR EMBL: AL389894; CAC22679.1;
 DR InterPro: IPR001365; A.deaminase.
 DR Pfam: PF00962; A.deaminase; 1.
 SQ SEQUENCE 2093 AA: 225856 MW: 40568264A9B7234D CRC64;

Query Match 89.5%; Score 34; DB 5; Length 2093;
 Best Local Similarity 83.3%; Pred. No. 4,9e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FKPLMR 6
 DB 1535 FEPLMR 1540

RESULT 11
 O9V7H5 PRELIMINARY: PRT: 164 AA.

AC O9V7H5;
 DT 01-MAY-2000 (TEMBLrel. 13, Created)
 DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TEMBLrel. 13, Last annotation update)
 DE CG8386 PROTEIN.
 GN CG8386.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Baidew R.M., Basu A., Baxendale J., Bayraktaroglu H., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burkova D., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrelia S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iltisgram C.,
 RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kennison J.P., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
 RA Palazozo M., Peltman G.S., Pan S., Pollard J., Puti V., Reese M.C.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,

RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yel R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003808; AAF58080.1; -
 DR FlyBase: FBgn0034061; CC8386.
 SO SEQUENCE 167 AA; 19004 MW; 24D45FAFF25D3EDA CRC64;

Query Match 86.8%; Score 33; DB 5; Length 164;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FKPLM 5
 |||||
 Db 121 FKPLM 125

RESULT 12
 ID 09P009 PRELIMINARY; PRT: 167 AA.
 AC 09P009;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE HSPC15.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Blood;
 RA Ye M., Zhang Q.H., Zhou J., Shen Y., Wu X.Y., Guan Z.Q., Wang L.,
 RA Fan H.Y., Mo Y.F., Dai M., Huang Q.H., Chen S.J., Chen Z.;
 RT "Human full length cDNA cloned from cd14+ stem cells.";
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF161504; AAF29119.1; -
 SO SEQUENCE 167 AA; 19467 MW; 3875D9187DC43E03 CRC64;

Query Match 86.8%; Score 33; DB 4; Length 167;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FKPLM 5
 |||||
 Db 121 FKPLM 125

RESULT 13
 ID 09BS96 PRELIMINARY; PRT: 167 AA.
 AC 09BS96;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE SIMILAR TO HYPOTHETICAL PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BLADDER CARCINOMA;
 RA Strausberg R.;
 RT Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC005187; AA005187.1; -

SO SEQUENCE 167 AA; 19398 MW; 06BF63787DD8D0E6 CRC64;

Query Match 86.8%; Score 33; DB 4; Length 167;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FKPLM 5
 |||||
 Db 121 FKPLM 125

RESULT 14
 ID 09CR09 PRELIMINARY; PRT: 167 AA.
 AC 09CR09;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE 1110021H02RIK PROTEIN.
 GN 1110021H02RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRATIN-C57BL/6J; TISSUE=EMBRYO, AND SPLEEN;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamane K.,
 RA Saito T., Okazaki Y., Cojocari T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gliss C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schirral L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carrinci P., de Ronaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Garlond M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriquez I., Sakamoto N.,
 PA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Wetz C., Whitaker C., Wilmink L.,
 RA Wyshak-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL: AK003897; BAB23063.1; -
 DR EMBL: AK003070; BAB22546.1; -
 DR MGI: 1913405; 1110021H02RIK.
 SO SEQUENCE 167 AA; 19481 MW; 359C44F2CE1FE3E0 CRC64;

Query Match 86.8%; Score 33; DB 11; Length 167;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FKPLM 5
 |||||
 Db 121 FKPLM 125

RESULT 15
 ID 09SXC8 PRELIMINARY; PRT: 174 AA.
 AC 09SXC8;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE T17H3.3 PROTEIN.
 GN T17H3.3.

OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae:
OC eurosids II: Brassicales: Brassicaceae: Arabidopsis.
OX NCBI_taxid=3702;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA:
RA Vysotskaia V.S., Schwartz J.R., Yu G., Tortum M., Lenz C., Liu S.,
RA Lee J.M., Liu A., Li J., Gonzalez A., Liu K., Vaysberg M., Sakano H.,
RA Chin C., Howng B., Choi E., Chou H., Altai H., Araujo R., Brooks S.,
RA Buehler E., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,
RA Halzar L., Khan S., Kim C., Palm C., Rowley D., Shinn P., Walker M.,
RA Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
RT "Arabidopsis thaliana chromosome 1 BAC T17H3 sequence."
RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC005916; AAD45991.L; -
SQ SEQUENCE 174 AA; 19611 MW; AF17805A92D9ED6 CRC64;

Query Match 86.8%; Score 33; DB 10; Length 174;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FKPLM 5
11111
DB 124 FKPLM 128

Search completed: February 27, 2002, 11:50:11
Job time: 990 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:36:04 ; Search time 132.19 Seconds
(without alignments)
1.021 Million cell updates/sec

Title: US-09-446-109A-10
Perfect score: 38
Sequence: 1 FKPLMR 6

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/2/laa/5A-COMB.pep.*
2: /cgn2_6/ptodata/2/laa/5B-COMB.pep.*
3: /cgn2_6/ptodata/2/laa/6A-COMB.pep.*
4: /cgn2_6/ptodata/2/laa/6B-COMB.pep.*
5: /cgn2_6/ptodata/2/laa/PCFUS-COMB.pep.*
6: /cgn2_6/ptodata/2/laa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	DB	ID	Description
1	33	86.8	6	1	US-08-215-137-13		Sequence 13, Appl
2	31	81.6	16	1	US-08-079-051-2		Sequence 2, Appl
3	31	81.6	115	3	PCT-US94-06994-2		Sequence 351, Appl
4	31	81.6	115	3	US-08-513-9748-351		Sequence 35, Appl
5	31	81.6	304	5	US-08-118-270-35		Sequence 35, Appl
6	31	81.6	350	2	PCT-US93-08528-35		Sequence 9, Appl
7	31	81.6	482	2	US-08-458-970A-9		Sequence 2, Appl
8	31	81.6	592	3	US-08-876-874-2		Sequence 2, Appl
9	31	78.9	592	3	US-08-991-813-2		Sequence 3, Appl
10	29	76.3	16	3	US-08-312-2028-3		Sequence 3, Appl
11	29	76.3	16	3	US-09-079-347-3		Sequence 3, Appl
12	29	76.3	16	4	US-08-075-725-3		Sequence 3, Appl
13	29	76.3	16	4	US-08-809-646-3		Sequence 3, Appl
14	29	76.3	16	5	PCT-US95-12433-3		Sequence 3, Appl
15	29	76.3	175	3	US-08-984-550-2		Sequence 2, Appl
16	29	76.3	175	3	US-09-103-359-14		Sequence 14, Appl
17	29	76.3	179	1	US-08-049-473-32		Sequence 32, Appl
18	29	76.3	179	1	US-08-312-648-32		Sequence 32, Appl
19	29	76.3	179	5	PCT-US94-04190-32		Sequence 32, Appl
20	29	76.3	180	1	US-08-049-473-31		Sequence 31, Appl
21	29	76.3	180	1	US-08-312-648-31		Sequence 31, Appl
22	29	76.3	180	5	PCT-US94-04190-31		Sequence 31, Appl
23	29	76.3	181	1	US-08-049-473-27		Sequence 27, Appl
24	29	76.3	181	1	US-08-049-473-28		Sequence 28, Appl
25	29	76.3	181	1	US-08-049-473-29		Sequence 29, Appl
26	29	76.3	181	1	US-08-312-648-27		Sequence 27, Appl
27	29	76.3	181	1	US-08-312-648-28		Sequence 28, Appl

ALIGNMENTS

28	29	76.3	181	1	US-08-312-648-29	Sequence 29, Appl
29	29	76.3	181	1	US-08-418-444A-3	Sequence 3, Appl
30	29	76.3	181	1	US-08-418-444A-4	Sequence 4, Appl
31	29	76.3	181	1	US-08-418-444A-5	Sequence 5, Appl
32	29	76.3	181	1	US-08-418-444A-6	Sequence 6, Appl
33	29	76.3	181	5	PCT-US94-04190-27	Sequence 27, Appl
34	29	76.3	181	5	PCT-US94-04190-28	Sequence 28, Appl
35	29	76.3	181	5	PCT-US94-04190-29	Sequence 29, Appl
36	29	76.3	458	2	US-08-655-878-2	Sequence 2, Appl
37	29	76.3	479	2	US-08-899-514-2	Sequence 2, Appl
38	29	76.3	501	1	US-08-722-001-14	Sequence 14, Appl
39	29	76.3	501	2	US-08-467-568-9	Sequence 9, Appl
40	29	76.3	501	2	US-09-030-582-9	Sequence 9, Appl
41	29	76.3	521	2	US-08-406-855A-19	Sequence 19, Appl
42	29	76.3	521	3	US-09-206-899-19	Sequence 19, Appl
43	29	76.3	572	1	US-08-334-698-2	Sequence 2, Appl
44	29	76.3	572	1	US-08-228-932-2	Sequence 2, Appl
45	29	76.3	572	1	US-08-468-939-2	Sequence 2, Appl

RESULT 1
US-08-215-137-13
Sequence 13, Application US/08215137
Patent No. 5614370
GENERAL INFORMATION:
APPLICANT: Konleat's, Zenon
APPLICANT: Siciliano, Salvatore J
APPLICANT: Springer, Martin S
TITLE OF INVENTION: ASSAY TO IDENTIFY HUMAN C5A ANTAGONISTS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESS: Merck & Co., Inc.
STREET: 126 Lincoln Ave., P.O. Box 2000
CITY: Rahway
STATE: New Jersey
COUNTRY: United States of America
Zip: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/215,137
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Benzen, Gerard H
REGISTRATION NUMBER: 35,746
REFERENCE/DOCKET NUMBER: 19108
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-3901
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 6 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /label-BH
OTHER INFORMATION: "either the natural phenylalanine amino
terminus or the Bolton-Hunter modified peptide

OTHER INFORMATION: having the group 3-(p-hydroxyphenyl)propionyl group

FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /label= dcha
OTHER INFORMATION: /note= "D-cyclohexylalanine"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 6
OTHER INFORMATION: /label= darg
OTHER INFORMATION: /note= "D-arginine"
US-08-215-137-13

Query Match 86.8%: Score 33; DB 1; Length 6;
Best Local Similarity 83.3%: Pred. No. 1.6e+05;
Matches 5: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPLMR 6
11111
Db 1 FKPKMR 6

RESULT 2
US-08-079-051-2
Sequence 2, Application US/08079051
Patent No. 5480974
GENERAL INFORMATION:
APPLICANT: MORGAN, EDWARD L.
APPLICANT: EMBER, JULIA A.
APPLICANT: HUGLI, TONY E.
TITLE OF INVENTION: ANTIBODIES TO HUMAN C5a RECEPTOR
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Spensley Horn Jubas & Lubitz
STREET: 1880 Century Park East, Suite 500
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/079,051
FILING DATE: 18-JUN-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Wetherell, Jr., Ph.D., John W.
REGISTRATION NUMBER: 31,678
REFERENCE/DOCKET NUMBER: PD-2790
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
CLONE: C5aR(133-148)
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..16
US-08-079-051-2

Query Match 81.6%: Score 31; DB 1; Length 16;
Best Local Similarity 80.0%: Pred. No. 4.2;

Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FKPLM 5
11111
Db 7 FKPLM 11

RESULT 3
PCT-US94-06994-2
Sequence 2, Application PC/TUS9406994
GENERAL INFORMATION:
APPLICANT: The Scripps Research Institute
TITLE OF INVENTION: ANTIBODIES TO HUMAN C5a RECEPTOR
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Spensley Horn Jubas & Lubitz
STREET: 1880 Century Park East, Suite 500
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/06994
FILING DATE: 20-JUN-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Tumarkin, Lisa A.
REGISTRATION NUMBER: P-38,347
REFERENCE/DOCKET NUMBER: PD-2790.
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
CLONE: C5aR(133-148)
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..16
PCT-US94-06994-2

Query Match 81.6%: Score 31; DB 5; Length 16;
Best Local Similarity 80.0%: Pred. No. 4.2;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FKPLM 5
11111
Db 7 FKPLM 11

RESULT 4
US-08-513-974B-351
Sequence 351, Application US/08513974B
Patent No. 6114139
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
APPLICANT: Hosoya, Masaki
APPLICANT: Fujii, Ryo
APPLICANT: Ohtaki, Tetsuya
APPLICANT: Fukusumi, Shoji
APPLICANT: Ohgi, Kazuhiko
TITLE OF INVENTION: G PROTEIN COUPLED RECEPTOR PROTEIN.

TITLE OF INVENTION: PRODUCTION, AND USE THEREOF
NUMBER OF SEQUENCES: 380
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/513,974B
FILING DATE: 14-SEP-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP95/01599
FILING DATE: 10-AUG-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-093989
FILING DATE: 19-AUG-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-057186
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-007177
FILING DATE: 20-JAN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-326611
FILING DATE: 28-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-270017
FILING DATE: 02-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-236357
FILING DATE: 30-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-236356
FILING DATE: 30-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-189274
FILING DATE: 11-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-189273
FILING DATE: 11-AUG-1945
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-189272
FILING DATE: 11-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Resnick, David S.
REGISTRATION NUMBER: 34,235
REFERENCE/DOCKET NUMBER: 45753
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 351:
SEQUENCE CHARACTERISTICS:
LENGTH: 115 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-513-974B-351

Query Match 81.6%; Score 31; DB 3; Length 115;
Best Local Similarity 80.0%; Pred. No. 32;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FKPLM 5

Db 4 FKPLM 8

RESULT 5
US-08-118-270-35
Sequence 35, Application US/08118270
Patent No. 5508384
GENERAL INFORMATION:
APPLICANT: Murphy, Randall B.
APPLICANT: Schuster, David I.
TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
NUMBER OF SEQUENCES: 348
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/118,270
FILING DATE: 09-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/943,236
FILING DATE: 10-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Townsend, Kevin G.
REGISTRATION NUMBER: 34,033
REFERENCE/DOCKET NUMBER: MURPHY-2A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248613
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 304 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-118-270-35

Query Match 81.6%; Score 31; DB 1; Length 304;
Best Local Similarity 80.0%; Pred. No. 86;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FKPLM 5

Db 103 FKPLM 107

RESULT 6
PCT-US93-08528-35
Sequence 35, Application PC/TUS9308528
GENERAL INFORMATION:
APPLICANT: New York University
TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
NUMBER OF SEQUENCES: 348
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA

ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/08528
FILING DATE: 09-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/943,236
FILING DATE: 10-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Townsend, Kevin G.
REGISTRATION NUMBER: 34,033
REFERENCE/DOCKET NUMBER: MURPHY-2 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ. ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 304 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US93-08528-35

Query Match 81.6%; Score 31; DB 5; Length 304;
Best Local Similarity 80.0%; Pred. No. 86;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FKPLW 5
Db 103 FKPLW 107

RESULT 7
US-08-458-970A-9
Sequence 9, Application US/08458970A
Patent No. 5861272
GENERAL INFORMATION:
APPLICANT: LI, ET AL.
TITLE OF INVENTION: C5a Receptor
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: CECCHI, BYRNE, BAIN, GILFILLAN,
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,970A
FILING DATE: June 2, 1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/09234
FILING DATE: 16 AUG 1994
ATTORNEY/AGENT INFORMATION:
NAME: MULLINS, J.C.
REGISTRATION NUMBER: 33,073
REFERENCE/DOCKET NUMBER: 325800-353
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700

TELEFAX: 201-994-1744
INFORMATION FOR SEQ. ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 350 amino acids
TYPE: amino acid
STRANDEDNESS: linear
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-458-970A-9

Query Match 81.6%; Score 31; DB 2; Length 350;
Best Local Similarity 80.0%; Pred. No. 1e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FKPLW 5
Db 139 FKPLW 143

RESULT 8
US-08-876-874-2
Sequence 2, Application US/08876874
Patent No. 5942405
GENERAL INFORMATION:
APPLICANT: Ames, Robert
APPLICANT: Bergsma, Dirk
APPLICANT: Foley, James
APPLICANT: Kumar, Chandrika
TITLE OF INVENTION: THERAPEUTIC AND SCREENING
TITLE OF INVENTION: METHODS USING C3A RECEPTOR AND C3A
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: RATNER & PRESTIA
STREET: P.O. BOX 980
CITY: VALLEY FORGE
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: PASTISO for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,874
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/019,627
FILING DATE: 16-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: PRESTIA, PAUL F.
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: P50501
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ. ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 482 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-876-874-2

Query Match 81.6%; Score 31; DB 2; Length 482;
Best Local Similarity 80.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 FKPLM 5
Db 125 FKPLM 129

RESULT 9
US-08-991-813-2
Sequence 2, Application US/08991813
Patent No. 6090579
GENERAL INFORMATION:
APPLICANT: ALBONE, EARL
APPLICANT: KIKLY, KRISTINE
TITLE OF INVENTION: HUMAN SDR2 CDNA CLONE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: RATNER & PRESTIA
STREET: P.O. BOX 980
CITY: VALLEY FORGE
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991.813
FILING DATE: 16-DEC-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/055,375
FILING DATE: 12-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: PRESTIA, PAUL F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GH-70281
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 592 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-991-813-2

Query Match 78.9% Score 30; DB 3; Length 592;
Best Local Similarity 80.0%; Pred. No. 2.7e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Caps 0;
OY 1 FKPLM 5
Db 395 FKPLM 399

RESULT 10
US-08-312-202B-3
Sequence 3, Application US/08312202B
Patent No. 5976816
Patent No. 5976816 5776705
GENERAL INFORMATION:
APPLICANT: ALKON, DANIEL L.;
APPLICANT: ETCHERRIGARAY, RENE; KIM, CHRISTOPHER S.;
APPLICANT: HAN, YI-FAN; NELSON, THOMAS J.
TITLE OF INVENTION: CELL TESTS FOR
TITLE OF INVENTION: ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/312.202B
FILING DATE: 26-SEP-1994
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/056,456
FILING DATE: 03-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: CAROL M. GRUPE
REGISTRATION NUMBER: 37,341
REFERENCE/DOCKET NUMBER: 2026-4075051
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 16
TYPE: AMINO ACID
STRANDEDNESS: UNKNOWN
TOPOLOGY: UNKNOWN
US-08-312-202B-3

Query Match 76.3% Score 29; DB 2; Length 16;
Best Local Similarity 80.0%; Pred. No. 10;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Caps 0;
OY 2 KPLMR 6
Db 1 RPLMR 5

RESULT 11
US-09-079-347-3
Sequence 3, Application US/09079347
Patent No. 6080582
GENERAL INFORMATION:
APPLICANT: ALKON, DANIEL L.;
APPLICANT: ETCHERRIGARAY, RENE; KIM, CHRISTOPHER S.;
APPLICANT: HAN, YI-FAN; NELSON, THOMAS J.
TITLE OF INVENTION: CELL TESTS FOR
TITLE OF INVENTION: ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/079,347
FILING DATE: 15-MAY-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/312,202

FILING DATE: 26-SEP-1994
PRIOR APPLICATION DATA: 08/056,456
FILING DATE: 03-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4075US3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 751-6849
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 16
TYPE: AMINO ACID
STRANDEDNESS: UNKNOWN
TOPOLOGY: UNKNOWN
US-09-079-347-3

Query Match 76.3%; Score 29; DB 3; Length 16;
Best Local Similarity 80.0%; Pred. No. 10;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 KPLMR 6
Db 1 RPLMR 5

RESULT 12
US-09-075-725-3
Sequence 3, Application US/09075725
Patent No. 6107050
GENERAL INFORMATION:
APPLICANT: ALKON, DANIEL L.;
APPLICANT: FAVIT, ANTONELLA; GRIMALDI, MAURIZIO;
APPLICANT: ELBAUM, DANEC; AND ZOHAR, OFER.
TITLE OF INVENTION: CELL TESTS AND DIAGNOSTIC
TITLE OF INVENTION: INDEX FOR ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/075,725
FILING DATE: 11-MAY-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/809,646
FILING DATE: 18-JULY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/12433
FILING DATE: 26-SEP-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/312,202
FILING DATE: 26-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/056,456
FILING DATE: 03-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4075US2
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 16
TYPE: AMINO ACID
STRANDEDNESS: UNKNOWN
TOPOLOGY: UNKNOWN
US-09-073-725-3

Query Match 76.3%; Score 29; DB 3; Length 16;
Best Local Similarity 80.0%; Pred. No. 10;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 KPLMR 6
Db 1 RPLMR 5

RESULT 13
US-08-809-646-3
Sequence 3, Application US/08809646
Patent No. 6300085
GENERAL INFORMATION:
APPLICANT: ALKON, DANIEL L.; ETCHEBERRICARAY,
APPLICANT: RENE; KIM, CHRISTOPHER S.; HAN,
APPLICANT: YI-FAN; AND NELSON, TOM J.
TITLE OF INVENTION: CELL TESTS AND DIAGNOSTIC
TITLE OF INVENTION: INDEX FOR ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/809,646
FILING DATE: 18-JULY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/12433
FILING DATE: 26-SEP-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/312,202
FILING DATE: 26-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/056,456
FILING DATE: 03-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: RICHARD W. BORK
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4075US2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 16
TYPE: AMINO ACID
STRANDEDNESS: UNKNOWN
TOPOLOGY: UNKNOWN
US-08-809-646-3

Query Match 76.3% Score 29; DB 4; Length 16;
Best Local Similarity 80.0% Pred. No. 10;
Matches 4: Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 KPLMR 6
:1111
DB 1 RPLMR 5

RESULT 14
PCT-US95-12433-3
Sequence 3: Application PC/TUS9512433
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: CELL TESTS AND DIAGNOSTIC
TITLE OF INVENTION: INDEX FOR ALZHEIMER'S DISEASE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/12433
FILING DATE: 26-SEP-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/312,202
FILING DATE: 26-SEP-1994
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/056,456
FILING DATE: 03-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: WILLIAM S. FEILER
REGISTRATION NUMBER: 26,728
REFERENCE/DOCKET NUMBER: 2026-4075PCTIII
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 16
TYPE: AMINO ACID
STRANDEDNESS: UNKNOWN
TOPOLOGY: UNKNOWN
PCT-US95-12433-3

Query Match 76.3% Score 29; DB 5; Length 16;
Best Local Similarity 80.0% Pred. No. 10;
Matches 4: Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 KPLMR 6
:1111
DB 1 RPLMR 5

RESULT 15
US-08-984-550-2
Sequence 2: Application US/08984550
Patent No. 6054639
GENERAL INFORMATION:
APPLICANT: MCGONIGLE, BRIAN
APPLICANT: O'KEEFE, DANIEL

TITLE OF INVENTION: SOYBEAN ADP RIBOSYLATION FACTOR
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: U.S.A.
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.50 INCH
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WINDOWS 95
SOFTWARE: MICROSOFT WORD VERSION 7.0A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/984,550
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: FLOYD, LINDA AXAMETHY
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: CL-1109
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-992-8112
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 175 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHETICAL: YES
AMT-SENSE: NO
US-08-984-550-2

Query Match 76.3% Score 29; DB 3; Length 175;
Best Local Similarity 80.0% Pred. No. 1,2e+02;
Matches 4: Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 KPLMR 6
:1111
DB 69 RPLMR 73

Search completed: February 27, 2002, 11:36:05
Job time: 144 sec

us-09-446-109a-11.rag

OM protein - protein search, using sw model

1.464 Million cell updates/sec

Scoring table: BLOSUM62

Total number of hits satisfying chosen parameters: 522463

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1	/SID8/gcgdata/gemseq/gemseqp/AA1980.DAT *
2	/SID8/gcgdata/gemseq/gemseqp/AA1981.DAT *
3	/SID8/gcgdata/gemseq/gemseqp/AA1982.DAT *
4	/SID8/gcgdata/gemseq/gemseqp/AA1983.DAT *
5	/SID8/gcgdata/gemseq/gemseqp/AA1984.DAT *
6	/SID8/gcgdata/gemseq/gemseqp/AA1985.DAT *
7	/SID8/gcgdata/gemseq/gemseqp/AA1986.DAT *
8	/SID8/gcgdata/gemseq/gemseqp/AA1987.DAT *
9	/SID8/gcgdata/gemseq/gemseqp/AA1988.DAT *
10	/SID8/gcgdata/gemseq/gemseqp/AA1989.DAT *
11	/SID8/gcgdata/gemseq/gemseqp/AA1990.DAT *
12	/SID8/gcgdata/gemseq/gemseqp/AA1991.DAT *
13	/SID8/gcgdata/gemseq/gemseqp/AA1992.DAT *
14	/SID8/gcgdata/gemseq/gemseqp/AA1993.DAT *
15	/SID8/gcgdata/gemseq/gemseqp/AA1994.DAT *
16	/SID8/gcgdata/gemseq/gemseqp/AA1995.DAT *
17	/SID8/gcgdata/gemseq/gemseqp/AA1996.DAT *
18	/SID8/gcgdata/gemseq/gemseqp/AA1997.DAT *
19	/SID8/gcgdata/gemseq/gemseqp/AA1998.DAT *
20	/SID8/gcgdata/gemseq/gemseqp/AA1999.DAT *
21	/SID8/gcgdata/gemseq/gemseqp/AA2000.DAT *
22	/SID8/gcgdata/gemseq/gemseqp/AA2001.DAT *

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
33	94.3	6	13	AA830577	Anaphylatoxin ant	
33	94.3	6	16	AA889737	Csa peptide analog	
30	85.7	39	20	AAW74031	human D2H binding	
30	85.7	434	21	AA801208	Corn putative leci	
7	85.7	462	22	AA898875	E. coli growth and	
	85.7	700	21	AAV58359	Yeast G protein-co	
	85.7	847	21	AAV58360	Yeast G protein-co	
	85.7	864	21	AAV58358	Yeast G protein-co	
	5.7	953	21	AAV58356	Yeast G protein-co	
	7	953	21	AAV58357	Yeast G protein-co	
	7	961	21	AAV58355	Yeast G protein-co	

12	30	35.7	1093	14	AAR41001
13	30	85.7	2204	11	AAV51233
14	29	82.9	90	21	AAB53988
15	29	82.9	217	22	AAB93531
16	29	82.9	462	22	AAB72311
17	29	82.9	745	22	AAB673.3
18	28	80.0	13	19	AAV24555
19	28	80.0	13	19	AAV66337
20	28	80.0	13	19	AAW66337
21	28	80.0	13	19	AAW66337
22	28	80.0	13	21	AAV91816
23	28	80.0	13	21	AAV91816
24	28	80.0	13	21	AAV91855
25	28	80.0	18	16	AAW64899
26	28	80.0	19	14	AAR37902
27	28	80.0	54	22	AAW21765
28	28	80.0	54	22	AAW38099
29	28	80.0	65	21	AAW56337
30	28	80.0	65	21	AAW60389
31	28	80.0	71	22	AAW16822
32	28	80.0	71	22	AAW29313
33	28	80.0	71	22	AAW04533
34	28	80.0	82	22	AAW63599
35	28	80.0	90	21	AAW19055
36	28	80.0	94	21	AAW01365
37	28	80.0	107	21	AAW40665
38	28	80.0	108	21	AAW21122
39	28	80.0	111	21	AAB40474
40	28	80.0	117	21	AAB26788
41	28	80.0	120	21	AAW26044
42	28	80.0	120	22	AAW02544
43	28	80.0	126	21	AAW03611
44	28	80.0	128	21	AAW10411
45	28	80.0	128	21	AAW40665

ALIGNMENTS

RESULT	1
AAAR30577	
7D	AAAR30577 standard; peptide; 6 AA.
AC	
AAAR30577;	
29-JAN-1993	(first entry)
DT	
XX	
DE	Anaphylatoxin antagonist or agonist peptide.
XX	
XX	Anaphylatoxin; receptor; ligand; antiinflammatory; hexapeptide;
KM	hepaleptide; immunodeficiency; allergy; autoimmune; ARDS; cancer
KM	infection; endotoxin; asthma; gout; psoriasis; cirrhosis;
KM	inflammation; bowel; disease; hepatitis; burns; myocardial;
KM	infection; transplant rejection; ischemic.
XX	
OS	Synthetic.
XX	
XX	
FT	key
FT	Modified-site
FT	1 location/Qualifiers
FT	/note= "N-Me-Phe"
FT	4
FT	/label= OTHER
FT	/note= "(2R)-2-amino-3-cyclohexyl-propanoyl"
FT	6
FT	/note= "D-Arg"
XX	
XX	
IN	MO9211858-A.
XX	
AD	23-JUL-1992.
XX	
FE	10-DEC-1991.
XX	
XX	91WO-US09319.
XX	
XX	90US-0634641.
XX	

XX (ABBO) ABBOTT LABORATORIES.
 XX
 XX
 PI Kewal M, Luly JR, Or YS, Wagner R, Wiedeman PE;
 XX
 DR WPI: 1992-268383/32.
 XX
 PT New hexa- and hepta-peptide(s) are anaphylatoxin antagonists and
 PT agonists - for treating inflammatory and immunodeficiency
 PT diseases, cancers and severe infections
 XX
 PS Claim 11: Page 153; 160pp; English.
 XX
 CC The peptide is a specifically claimed example of a group of highly
 CC generic hexa- and heptapeptides which are (a) anaphylatoxin
 CC antagonists useful for treating asthma, other allergies,
 CC inflammations, autoimmune diseases, serum sickness, gout, bullous
 CC skin diseases, psoriasis, ARDS, endotoxin shock, hepatic cirrhosis,
 CC pancreatitis, inflammatory bowel disease, burns, sepsis, myocardial
 CC infarction, chronic hepatitis, transplant rejection, or ischaemic
 CC heart or brain damage; or (b) anaphylatoxin agonists useful for
 CC stimulating inflammatory and immune responses, e.g. in the treatment
 CC of cancer, immunodeficiency diseases and severe infections.
 XX
 SQ Sequence 6 AA:
 XX
 Query Match 94.3%; Score 33; DB 13; Length 6;
 Best Local Similarity 100.0%; Pred. No. 4.3e+05;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 FKPKWR 6
 111111
 Db 1 fkpwr 6
 XX
 RESULT 2
 AAR89737
 ID AAR89737 standard; peptide; 6 AA.
 XX
 AC AAR89737;
 XX
 DT 03-MAY-1996 (first entry)
 XX
 DE C5a peptide analogue, C089, used to identify C5a (ant)agonists.
 XX
 KW C5a: complement; agonist; antagonist; peptide; binding assay;
 KW identification; inflammation; pain reduction; respiratory disorder;
 KW cardiovascular; hypertension; ARDS; anaphylaxis; arthritis; asthma;
 KW competitive.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT MISC-difference 1 /label= Phe, OTHER
 FT /note= OTHER = the Bolton-Hunter modified peptide
 FT having the 3-(p-hydroxyphenyl)-propionyl
 FT group
 FT MISC-difference 4 /label= OTHER
 FT /note= "D-cyclohexylalanine"
 FT Modified-site 6 /label= OTHER
 FT /note= "cyclohexylalanine"
 FT
 XX MO9525957-A1.
 XX
 PC 28-SEP-1995.
 XX
 PF 14-MAR-1995; 95WO-US03209.
 XX
 PK 18-MAR-1994; 94US-0215137.

XX (MERI) MERCK & CO INC.
 XX
 XX
 PI Konteatis Z, Siciliano SJ, Springer MS;
 XX
 DR WPI: 1995-344718/44.
 XX
 PT C5a receptor binding assays - used for identifying cpds. having C5a
 PT antagonist, agonist or partial agonist activity
 XX
 PS Claim 15: Page 47; 65pp; English.
 XX
 CC C5a C-terminal peptide analogues AAR89734-44 and AAR90033-35 may be used
 CC in a new method for identifying C5a agonists/antagonists. AAR89734
 CC represents the generic formula of this newly defined class of
 CC labelled peptides. C5a antagonists identified using this method are
 CC useful in the treatment of a wide variety of C5a-mediated diseases
 CC such as acute respiratory distress syndrome (ARDS), anaphylactic
 CC shock, psoriasis, osteoarthritis, rheumatoid arthritis, asthma,
 CC cardiovascular disorders, and metastatic spread of cancerous tumours.
 XX
 SQ Sequence 6 AA:
 XX
 Query Match 94.3%; Score 33; DB 16; Length 6;
 Best Local Similarity 100.0%; Pred. No. 4.3e+05;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 FKPKWR 6
 111111
 Db 1 fkpwr 6
 XX
 RESULT 3
 AAW74031
 ID AAW74031 standard; peptide; 39 AA.
 XX
 AC AAW74031;
 XX
 DT 04-MAY-1999 (first entry)
 XX
 DE Human D2H binding protein DAB3.
 XX
 KW Gastro-intestinal transport receptor; binding protein; hsi; hPT1;
 KW D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
 KW intestinal peptide-associated transporter; hypertension; diabetes;
 KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KW therapeutic agent delivery; therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO9851325-A2.
 XX
 TD 19-NOV-1998.
 XX
 PF 15-MAY-1998; 98WO-US10088.
 XX
 PR 15-MAY-1997; 97US-0046595.
 XX
 PA (CYTO-) CYTOGEN CORP.
 PA (ELAN-) ELAN CORP PLC.
 XX
 PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambdin LJ;
 PI Omahony DJ, Patterson CA, Singleton J;
 XX
 DR WPI: 1999-009568/01.
 XX
 PT New proteins that bind specifically to receptors in the
 PT gastro-intestinal tract and related nucleic acid - chimaeras and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastro-intestinal tract, e.g. insulin or leuprolide
 XX
 PS Claim 2: Page 54; 294pp; English.

XX This sequence represents a peptide that specifically binds to the human
 CC D2H protein. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (HPI1),
 CC HPI1, D2H and human sucrose-isomaltase complex (hst). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents
 CC and antigens. (I) may also provide targeting to the GI tract. Other uses
 CC of (I) are: (i) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (ii) to screen for molecules that bind (I).
 CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.
 CC
 CC Sequence 39 AA:

Query Match 85.7% Score 30; DB 20; Length 39;
 Best Local Similarity 66.7% Pred. No. 18;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Caps 0;

OY 1 FKPMR 6
 1:1 11
 DB 15 frpwr 20

RESULT 4

AAB01208 standard: Protein: 434 AA.

AAB01208:

12-DEC-2000 (first entry)

Corr putative lecithin:cholesterol acyltransferase #3.

Corr: lecithin:cholesterol acyltransferase; phytosterol;
 phosphatidylcholine-sterol O-acyltransferase; heat shock; cold shock.

Zea mays.

MO200032791-A2.

08-JUN-2000.

02-DEC-1999; 99WO-US28586.

03-DEC-1998; 98US-0110782.

(DUPO) DU PONT DE NEMOURS & CO E. I.

Cahoon RE, Kinney AJ, Sakai H, Shen JB, Butler KH, Saylor JJ;

WPI: 2000-412337/35.

N-PSDB: AAA49203.

Polynucleotide encoding plant lecithin:cholesterol acyltransferase

enzyme useful for producing transgenic plants and for producing

antibodies specific to which is useful for screening cDNA expression

libraries

Claim 10: Page 40-41: 49pp: English.

The present sequence is a putative protein sequence of a corn

lecithin:cholesterol acyltransferase (also known as

phosphatidylcholine-sterol O-acyltransferase). This enzyme is found

associated with high-density lipoproteins and Apolipoprotein-AI and -D.

CC The gene and protein can be used to produce transgenic plants which ha..
 CC increased lipid metabolism and membrane fluidity, and therefore increased
 CC resistance to heat and/or cold shock, to alter the content of phytosterol
 CC or lecithin in grains and to identify potential herbicides.

Sequence 434 AA:

Query Match 85.7% Score 30; DB 21; Length 434;
 Best Local Similarity 66.7% Pred. No. 2,1e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Caps 0;

OY 1 FKPMR 6
 1:1 11
 DB 271 frpwr 276

RESULT 5

AAC98875 standard: Protein: 462 AA.

AAC98875:

26-SEP-2001 (first entry)

E. coli growth and proliferation related protein sequence SEQ ID NO:345.

Escherichia coli: growth; proliferation: microbial: antimicrobial;

bacterial infection; microorganism.

Escherichia coli.

MO200134810-A2.

17-MAY-2001.

09-NOV-2000; 2000WO-US30950.

09-NOV-1999; 99US-0164415.

(ELIT-) ELITRA PHARM INC.

Forsyth RA, Ohlsen K, Zyskind J;

WPI: 2001-335933/35.

N-PSDB: AAH84546.

Novel nucleic acids that inhibit Escherichia coli proliferation, useful

for screening for homologous genes and for designing expression vectors

Claim 19: Page 421-422: 522pp: English.

AAH84373 to AAH84499 represent Escherichia coli growth and proliferation

related DNA sequences (I). AAH84500 to AAH84670 encode the E. coli

growth and proliferation related proteins given in AAC99078 and AAC98830

to AAC98999. (I) can be used as potential targets for the generation of

new antimicrobial agents, and for identification of compounds which

interact with the gene products of (I). In addition the expression of

(I) and the purification of the proteins, the purified proteins can be

used to generate reagents and screen small molecule libraries or other

candidate compound libraries for compounds that can be further developed

to yield novel antimicrobial compounds. In addition, nucleic acid probes

complementary to (I) that are specific for particular species of

microorganisms can be used to identify particular microorganism species

in clinical specimens, therefore, providing a rapid and dependable

method by which to identify the causative agents of a bacterial

infection. Also, antibodies generated against proteins translated from

mRNA transcribed from proliferation-required sequences can also be used

to screen for specific microorganisms that produce such proteins in a

species-specific manner. AAH84371 and AAH84670 represent sequencing

primers used in the isolation of E. coli growth and proliferation

related sequence, which are used in an example from the present

CC Invention.
 XX
 SQ Sequence 462 AA;
 Query Match 85.7%; Score 30; DB 22; Length 462;
 Best Local Similarity 66.7%; Pred. No. 2.3e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FKPMR 6
 Db 398 ykplw.403
 RESULT 6
 ID AAY58359 standard; Protein: 700 AA.
 XX
 AC AAY58359;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE Yeast G protein-coupled receptor GPR1 mutant d694-954.
 XX
 KM Yeast: G protein-coupled receptor; GPR1; filamentous growth;
 KM pseudohyphal form; signalling pathway; antifungal; mutant; mutein.
 XX
 OS Saccharomyces cerevisiae.
 OS Synthetic.
 XX
 PN MO9963094-A2.
 XX
 PD 09-DEC-1999.
 XX
 PF 28-MAY-1999; 99MO-US11838.
 XX
 PR 01-JUN-1998; 98US-0088311.
 XX
 PA (MOUNT SINAI SCHOOL MEDICINE.
 PI Hirsch JP, Xue Y;
 DR WPI: 2000-086980/07.
 XX
 PT Novel protein, gene, antibody and recombinant cell useful for
 PT identifying modulators and inhibitors useful as anti-fungals -
 XX
 PS Example 6.2.5; Page -: 63pp; English.
 XX
 CC The invention relates to a novel yeast G protein-coupled receptor,
 CC GPR1 (AAY58355). GPR1 is predicted to have seven transmembrane domains,
 CC which is characteristic of G protein-coupled receptors, and also
 CC contains a very large third cytoplasmic loop and a large cytoplasmic
 CC tail. The third cytoplasmic loop contains two short basic sequences; one
 CC sequence (KRKRAQIC) is present at the N-terminal end of the loop and the
 CC other sequence (KKRRRAQIQ) is at the C-terminal end of the loop. The
 CC third cytoplasmic loop also contains an aspartagine-rich region. GPR1 is
 CC believed to provide the upstream signal that activates GPA2, and plays
 CC an essential role in inducing the switch from non-filamentous to
 CC filamentous growth in yeast. Modulators of GPR1 function, especially
 CC inhibitors, are useful for prevention or treatment of a disease or
 CC disorder involving a fungal infection. The inhibitor may be an anti-GPR1
 CC antibody. GPR1 antisense nucleic acid or a GPR1 gene altered so that an
 CC inserted heterologous sequence inactivates the biological activity of
 CC GPR1. Compounds which modulate GPR1 gene expression and/or gene product
 CC activity may be used to inhibit the conversion of a fungus from a
 CC non-filamentous form to a filamentous form. The present sequence
 CC represents a GPR1 mutant, d277-284, used in an exemplification of
 CC the invention. This sequence is lacking the majority of the cytoplasmic
 CC C-terminal tail, relative to the native GPR1.
 CC Note: This sequence is not shown in the specification, but is derived
 CC from the yeast GPR1 amino acid sequence given in figure 1A.
 CC
 XX

SQ Sequence 700 AA;
 Query Match 85.7%; Score 30; DB 21; Length 700;
 Best Local Similarity 66.7%; Pred. No. 3.5e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FKPMR 6
 Db 156 fkpwmk 161
 RESULT 7
 ID AAY58360 standard; Protein: 847 AA.
 XX
 AC AAY58360;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE Yeast G protein-coupled receptor GPR1 mutant d841-954.
 XX
 KM Yeast: G protein-coupled receptor; GPR1; filamentous growth;
 KM pseudohyphal form; signalling pathway; antifungal; mutant; mutein.
 XX
 OS Saccharomyces cerevisiae.
 OS Synthetic.
 XX
 PN MO9963094-A2.
 XX
 PD 09-DEC-1999.
 XX
 PF 28-MAY-1999; 99MO-US11838.
 XX
 PR 01-JUN-1998; 98US-0088311.
 XX
 PA (MOUNT SINAI SCHOOL MEDICINE.
 PI Hirsch JP, Xue Y;
 DR WPI: 2000-086980/07.
 XX
 PT Novel protein, gene, antibody and recombinant cell useful for
 PT identifying modulators and inhibitors useful as anti-fungals -
 XX
 PS Example 6.2.5; Page -: 63pp; English.
 XX
 CC The invention relates to a novel yeast G protein-coupled receptor,
 CC GPR1 (AAY58355). GPR1 is predicted to have seven transmembrane domains,
 CC which is characteristic of G protein-coupled receptors, and also
 CC contains a very large third cytoplasmic loop and a large cytoplasmic
 CC tail. The third cytoplasmic loop contains two short basic sequences; one
 CC sequence (KRKRAQIC) is present at the N-terminal end of the loop and the
 CC other sequence (KKRRRAQIQ) is at the C-terminal end of the loop. The
 CC third cytoplasmic loop also contains an aspartagine-rich region. GPR1 is
 CC believed to provide the upstream signal that activates GPA2, and plays
 CC an essential role in inducing the switch from non-filamentous to
 CC filamentous growth in yeast. Modulators of GPR1 function, especially
 CC inhibitors, are useful for prevention or treatment of a disease or
 CC disorder involving a fungal infection. The inhibitor may be an anti-GPR1
 CC antibody. GPR1 antisense nucleic acid or a GPR1 gene altered so that an
 CC inserted heterologous sequence inactivates the biological activity of
 CC GPR1. Compounds which modulate GPR1 gene expression and/or gene product
 CC activity may be used to inhibit the conversion of a fungus from a
 CC non-filamentous form to a filamentous form. The present sequence
 CC represents a GPR1 mutant, d277-284, used in an exemplification of
 CC the invention. This sequence is lacking a portion (designated the
 CC Smaest region) of the cytoplasmic C-terminal tail, relative to the
 CC native GPR1.
 CC Note: This sequence is not shown in the specification, but is derived
 CC from the yeast GPR1 amino acid sequence given in figure 1A.
 CC
 XX
 SQ Sequence 847 AA;
 XX

QY	1	FKPXR	6
			:
Db	156	fkpnwk	161

AC	AAV58358;
XX	
DT	27-MAR-2000 (first entry)

DE Yeast G protein-coupled receptor GPR1 mutant d490-586

KW Yeast; G protein-coupled receptor; GPRL, filamentous growth; pseudohyphal form; signalling pathway; mutant; mutein

OS Saccharomyces cerevisiae.
OS synthetic.

PN WO9963094-A2

PD 09-DEC-1999.

PF 28-MAY-1999; 99WO-US11838.

PR 01-JUN-1998; 9805-0088311.

PA (MOUN) MOUNT SINAI SCHOOL MEDICINE

PI Hirsch JP, Xue Y,

DR WPI; 2000-086980/07.

PT Novel protein, gene, antibody and recombinant cell useful for identifying modulators and inhibitors useful as anti-fungals -

PS Example 6.2.5; page -; 63pp; English

The invention relates to a novel yeast G protein-coupled receptor, GPR1 (AAV58355). GPR1 is predicted to have seven transmembrane domains, which is characteristic of G protein-coupled receptors and also contains a very large third cytoplasmic loop and a large cytoplasmic tail. The third cytoplasmic loop contains two short basic sequences; one sequence (KKRKAQIC) is present at the N-terminal end of the loop and the other sequence (KKRRRAQIQ) is at the C-terminal end of the loop. The third cytoplasmic loop also contains an asparagine-rich region. GPR1 is believed to provide the upstream signal that activates GPA2, and plays an essential role in inducing the switch from non-filamentous to filamentous growth in yeast. Modulators of GPR1 function, especially inhibitors, are useful for prevention or treatment of a disease or disorder involving a fungal infection. The inhibitor may be an anti-GPR1 antibody, GPR1 antisense nucleic acid or a GPR1 gene altered so that an inserted heterologous sequence inactivates the biological activity of GPR1. Compounds which modulate GPR1 gene expression and/or gene product activity may be used to inhibit the conversion of a fungus from a non-filamentous form to a filamentous form. The present sequence represents a GPR1 mutant, d277-284, used in an exemplification of the invention. This sequence is lacking the asparagine-rich region of the third cytoplasmic loop, relative to the native GPR1. Note: This sequence is not shown in the specification, but is derived from the yeast GPR1 amino acid sequence given in figure 1A.

Sequence	864 AA
...	

QY	1	FKPXR	6
			:
Db	156	fkpnw	161

PC AAY58356;
 XX
 DT 27-MAR-2000 (first entry)
 XX

DE Yeast G protein-coupled receptor GPR1 mutant d277-284
v.v.

KW Yeast; G protein-coupled receptor; GPR1; filamentous growth; pseudohyphal form; signalling pathway; mutant; mutein

05 *Saccharomyces cerevisiae*.
05 Synthetic.

PN WO9963094-A2

09-DEC-1999

PF 28-MAY-1999; 99WO-US11838.

01-JUN-1998; 9805-0088311.

PA (MOUN) MOUNT SINAI SCHOOL MEDICINE

PI Hirsch JP, Xue Y;

DR WPI; 2000-086490/07

Novel protein, gene, antibody and recombinant cell useful for

XX
PS
Example 6.2.5; Page -; 63pp; English.

CC The invention relates to a novel yeast

which is characteristic of G-protein-coupled receptors, and also contains a very large third cytoplasmic loop and a large cytoplasmic tail. The third cytoplasmic loop contains two short basic sequences; one sequence (KKRKADIG) is present at the N-terminal end of the loop and the other sequence (KKRRADYQ) is at the C-terminal end of the loop. The third cytoplasmic loop also contains an aspartate-rich region. GPR1 is believed to provide the upstream signal that activates G β γ 2, and plays an essential role in inducing the switch from non-filamentous to filamentous growth in yeast. Modulators of GPR1 function, especially inhibitors, are useful for prevention or treatment of a disease or disorder involving a fungal infection. The inhibitor may be an anti-GPR1 antibody, GPR1 antisense nucleic acid or a GPR1 gene altered so that an inserted heterologous sequence inactivates the biological activity of GPR1. Compounds which modulate GPR1 gene expression and/or gene product activity may be used to inhibit the conversion of a fungus from a non-filamentous form to a filamentous form. The present sequence represents a GPR1 mutant, d277-284, used in an exemplification of the invention. This sequence is lacking the first basic motif (KKRKADIG) of the third cytoplasmic loop, relative to the native GPR1. Note: This sequence is not shown in the specification, but is derived from the yeast GPR1 amino acid sequence given in figure 1A.

SQ Sequence 953 AA;

Query Match	85.7%;	Score 30;	DB 21;	length 9533;
Best Local Similarity	66.7%;	Pred. No. 4.7e+02;		

Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FKPMR 6
 ||| |:
 Db 156 fkpnrk 161

RESULT 10

AAV58357
 ID AAV58357 standard; Protein: 953 AA.

AC AAV58357;
 XX
 DT 27-MAR-2000 (first entry)

XX Yeast G protein-coupled receptor GPR1 mutant d610-617.

XX Yeast: G protein-coupled receptor; GPR1: filamentous growth;
 KM pseudohyphal form; signalling pathway: antifungal; mutant; mulein.

XX Saccharomyces cerevisiae.
 OS Synthetic.

XX WO9963094-A2.

XX 09-DEC-1999.

XX 28-MAY-1999; 99WO-US11838.

XX 01-JUN-1998; 98US-0088311.

XX (MOUN) MOUNT SINAI SCHOOL MEDICINE.

XX Hirsch JP, Xue Y;

XX WPI: 2000-086980/07.

PT Novel protein, gene, antibody and recombinant cell useful for
 PT identifying modulators and inhibitors useful as anti-fungals -

XX Example 6.2.5; Page -: 63pp; English.

XX The invention relates to a novel yeast G protein-coupled receptor,
 CC GPR1 (AAV58355). GPR1 is predicted to have seven transmembrane domains,
 CC which is characteristic of G protein-coupled receptors, and also
 CC contains a very large third cytoplasmic loop and a large cytoplasmic
 CC tail. The third cytoplasmic loop contains two short basic sequences: one
 CC sequence (KRKAQIG) is present at the N-terminal end of the loop and the
 CC other sequence (KKRAQIQ) is at the C-terminal end of the loop. The
 CC third cytoplasmic loop also contains an asparagine-rich region. GPR1 is
 CC believed to provide the upstream signal that activates GPA2, and plays
 CC an essential role in inducing the switch from non-filamentous to
 CC filamentous growth in yeast. Modulators of GPR1 function, especially
 CC inhibitors, are useful for prevention or treatment of a disease or
 CC disorder involving a fungal infection. The inhibitor may be an anti-GPR1
 CC antibody, GPR1 antisense nucleic acid or a GPR1 gene altered so that an
 CC inserted heterologous sequence inactivates the biological activity of
 CC GPR1. Compounds which modulate GPR1 gene expression and/or gene product
 CC activity may be used to inhibit the conversion of a fungus from a
 CC non-filamentous form to a filamentous form. The present sequence
 CC represents a GPR mutant, d610-617, used in an exemplification of
 CC the invention. This sequence is lacking the second basic motif
 CC (KKRAQIQ) of the third cytoplasmic loop, relative to the native GPR1.
 CC Note: This sequence is not shown in the specification, but is derived
 CC from the yeast GPR1 amino acid sequence given in figure 1A.

XX Sequence 953 AA;

Query Match 85.7%; Score 30; DB 21; Length 953;
 Best Local Similarity 66.7%; Pred. No. 4.7e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPMR 6
 ||| |:
 Db 156 fkpnrk 161

RESULT 11

AAV58355
 ID AAV58355 standard; Protein: 961 AA.

AC AAV58355;

XX 27-MAR-2000 (first entry)

XX Yeast G protein-coupled receptor GPR1.

XX Yeast: G protein-coupled receptor; GPR1: filamentous growth;
 KM pseudohyphal form; signalling pathway: antifungal.

XX Saccharomyces cerevisiae.

XX Location/Qualifiers

FT Region 1..55 /note= "Extracellular N-terminus"

FT Domain 56..80 /note= "Transmembrane domain 1"

FT Region 81..90 /note= "Intracellular loop 1"

FT Domain 91..115 /note= "Transmembrane domain 2"

FT Region 116..132 /note= "Extracellular loop 1"

FT Domain 133..156 /note= "Transmembrane domain 3"

FT Region 157..180 /note= "Intracellular loop 2"

FT Domain 181..200 /note= "Transmembrane domain 4"

FT Region 201..246 /note= "Extracellular loop 2"

FT Domain 247..276 /note= "Transmembrane domain 5"

FT Misc-difference 250 /note= "Encoded by AGC"

FT Region 277..620 /note= "Intracellular loop 3"

FT Region 277..284 /note= "Basic motif 1 (KRKAQIG)"

FT Region 490..586 /note= "Asparagine-rich region"

FT Region 610..617 /note= "Basic motif 2 (KKRAQIQ)"

FT Domain 621..645 /note= "Transmembrane domain 6"

FT Region 646..656 /note= "Extracellular loop 3"

FT Domain 657..678 /note= "Transmembrane domain 7"

FT Region 679..961 /note= "Intracellular C-terminus"

FT Misc-difference 250 /note= "Encoded by AGC"

FT WO9963094-A2.

XX 09-DEC-1999.

XX 28-MAY-1999; 99WO-US11838.

XX 01-JUN-1998; 98US-0088311.

XX (MOUN) MOUNT SINAI SCHOOL MEDICINE.

XX Hirsch JP, Xue Y;

PI

XX WPI: 2000-086980/07.
 DR N-PSDB; AA255699.
 XX
 PT Novel protein, gene, antibody and recombinant cell useful for
 XX identifying modulators and inhibitors useful as anti-fungals -
 PS
 XX Claim 4; Fig 1A; 63pp; English.
 CC This sequence represents a novel yeast G protein-coupled receptor,
 CC GPR1. GPR1 is predicted to have seven transmembrane domains,
 CC which is characteristic of G protein-coupled receptors, and also
 CC contains a very large third cytoplasmic loop and a large cytoplasmic
 CC tail. The third cytoplasmic loop contains two short basic sequences; one
 CC sequence (KRKRDIG) is present at the N-terminal end of the loop and the
 CC other sequence (KKRRADIG) is at the C-terminal end of the loop. The
 CC third cytoplasmic loop also contains an asparagine-rich region. GPR1 is
 CC believed to provide the upstream signal that activates GPa2, and plays
 CC an essential role in inducing the switch from non-filamentous to
 CC filamentous growth in yeast. Modulators of GPR1 function, especially
 CC inhibitors, are useful for prevention or treatment of a disease or
 CC disorder involving a fungal infection. The inhibitor may be an anti-GPR1
 CC antibody, GPR1 antisense nucleic acid or a GPR1 gene altered so that an
 CC inserted heterologous sequence inactivates the biological activity of
 CC GPR1. Compounds which modulate GPR1 gene expression and/or gene product
 CC activity may be used to inhibit the conversion of a fungus from a
 CC non-filamentous form to a filamentous form.
 CC
 SQ Sequence 961 AA:
 Query Match 85.7%; Score 30; DB 21; Length 961;
 Best Local Similarity 66.7%; Pred. No. 4.8e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FKRXWR 6
 III I:
 Db 156 fKpwnk 161
 RESULT 12
 AAR41001
 ID AAR41001 standard; Protein: 1093 AA.
 XX
 AC AAR41001:
 DT 25-FEB-1994 (first entry)
 XX
 DE Human myotonic dystrophy gene protein.
 XX
 KW Abnormality: muscular dystrophy; CHR 19; chromosome 19;
 XX protein kinase; polymerase chain reaction; brain.
 OS
 XX Homo sapiens.
 FH
 FT Key Location/Qualifiers
 FT 1..1093 /note="encoded by predicted reading frame a,
 FT x's in the sequence indicate stop codons
 FT in the reading frame"
 XX
 PN MO9317104-A.
 XX
 PD 02-SEP-1993.
 XX
 PF 19-FEB-1993; 93WO-US01545.
 XX
 PR 20-FEB-1992; 92US-0839255.
 XX
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 XX
 PI Hbrook JD, Housman DE;
 XX

DR WPI: 1993-288410/36.
 XX
 PT DNA sequence of myotonic dystrophy gene - used to produce probes
 PT and identify CHR 19 abnormality and protein kinase responsible
 XX
 PS Disclosure; Fig 6; 64pp; English.
 XX
 CC The sequence is that encoded by predicted reading frame a of
 CC the human myotonic dystrophy (DM) gene. It may be used in the
 CC identification of individuals affected by DM.
 CC
 SQ Sequence 1093 AA:
 Query Match 85.7%; Score 30; DB 14; Length 1093;
 Best Local Similarity 66.7%; Pred. No. 5.5e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FKRXWR 6
 III I:
 Nb 382 fKpwnr 387
 RESULT 13
 AAV51233
 ID AAV51233 standard; Protein: 2204 AA.
 XX
 AC AAV51233:
 DT 07-APR-2000 (first entry)
 XX
 DE Newcastle disease virus Lasota genome encoded protein 6.
 XX
 DE Newcastle disease virus; Infection; lentogenic; F protein; vaccine;
 KW respiratory disease; gastrointestinal disease; poultry pathogen;
 KW local immunity.
 XX
 CS Newcastle disease virus.
 XX
 OS
 XX MO9966045-A1.
 PN
 PD 23-DEC-1999.
 XX
 PP 17-JUN-1999; 99WO-NL00377.
 XX
 FR 19-JUN-1998; 98EP-0202054.
 XX
 PA (DIEN-) STICHTING DIENST LANDBOUWKUNDIG ONDERZOE.
 XX
 PI Peeters BPH, De Leeuw OS, Koch G, Gielkens ALJ;
 XX
 DR WPI: 2000-106102/09.
 XX
 PT New avian paramyxovirus cDNA, useful for production of vaccine against
 PT Newcastle disease virus
 XX
 PS Disclosure; Fig 3; 115pp; English.
 XX
 CC This invention describes a novel avian paramyxovirus cDNA (I) which
 CC comprises a nucleic acid sequence corresponding to the 5' terminal
 CC end of the genome of avian-paramyxovirus allowing the generation of
 CC an infectious copy of avian-paramyxovirus. The cell line is useful for
 CC the production of infectious lentogenic NDV (Newcastle Disease virus)
 CC without the addition of exogenous proteolytic activity. Also it is
 CC possible to generate a stable transfected cell line that expresses the
 CC wild-type F protein in the virus envelope therefore providing infectious
 CC particles, useful in the form of a vaccine, especially against
 CC respiratory and/or gastrointestinal diseases. NDV can be easily cultured
 CC to very high titers in embryonated eggs. Mass culture of embryonated
 CC eggs is relatively cheap. NDV vaccines are relatively stable and can be
 CC simply administered by mass application methods e.g. drinking water or
 CC by spraying or by aerosol formation. The natural route of infection is
 CC by the respiratory and/or gastrointestinal tract which are also the major

CC routes of infection of many other poultry pathogens. NDV can induce local
 CC immunity despite the presence of circulating maternal antibody. This
 CC sequence represents a protein encoded by the NDV strain LaSota genome
 CC which is described in the method of the invention.

XX Sequence 2204 AA:

Query Match 85.7%; Score 30; DB 21; Length 2204;
 Best Local Similarity 66.7%; Pred. No. 1.1e+03;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR-6
 11111
 Db 1844 frplwr 1849

RESULT 14

AAB53989
 ID AAB53989 standard; Protein; 90 AA.

AC AAB53989;

DT 09-MAR-2001 (first entry)

DE Human colon cancer antigen protein sequence SEQ ID NO:1529.

XX Human: colon cancer: colon cancer antigen: diagnosis: detection;
 KW Identification; cytostatic; cardioactive; neuroprotective; vulnary;
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;
 KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
 KW neural disorder; immune system disorder; muscular disorder;
 KW reproductive disorder; gastrointestinal disorder; renal disorder;
 KW Infectious disease; cardiovascular disorder.

XX Homo sapiens.

PN W020005531-A1.

XX 21-SEP-2000.

PF 08-MAR-2000; 2000MO-US05883.

PR 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

PA Rosen CA, Ruben SM;

DR WPI: 2000-587534/55.

XX N-PSDB: AAC98746.

PT Colon cancer associated gene sequences, referred to as colon cancer
 PT antigens, useful for the treatment, prevention, and diagnosis of colon
 PT disorders such as colon cancer -

PS Claim 11: Page 2083-2084; 2104pp: English.

XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
 CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
 CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
 CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
 CC vulnary, nephrotropic, antiinfective and antibacterial activities, and
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,
 CC proteins and antibodies to the proteins are useful for the prevention,
 CC treatment and diagnosis of colon disorders, such as colon cancer. The
 CC polynucleotides may be used in diagnostics and research, such as for
 CC chromosome identification, and as hybridisation probes. The proteins
 CC may also be used to prevent diseases such as neural disorders, immune
 CC system disorders, muscular disorders, reproductive disorders,
 CC gastrointestinal disorders, wounds, renal disorders, infectious
 CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
 CC AAB54007 represent sequences used in the exemplification of the present

CC invention.

XX Sequence 90 AA:

Query Match 82.9%; Score 29; DB 21; Length 90;
 Best Local Similarity 66.7%; Pred. No. 67;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
 11111
 Db 74 fkpckw 79

RESULT 15

AAB93531
 ID AAB93531 standard; Protein; 217 AA.

AC AAB93531;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:12886.

XX Human: primer: detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

PN EP1074617-A2.

XX 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI: 2001-318749/34.

PS Claim 8: SEQ ID 12886; 2537pp + CD ROM: English.

XX The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

Wed Feb 27 12:13:44 2002

us-09-446-109a-11.rag

CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.

SO Sequence 217 AA;

Query Match 82.9%; Score 29; DB 22; Length 217;

Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPM 5.
114-1
Db 164 fkpwm 168

Search completed: February 27, 2002, 11:41:17
Job time: 456 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:45:20 : Search time 145.23 Seconds
(without alignments)
3.147 Million cell updates/sec

Title: US-09-446-109A-11

Perfect score: 35

Sequence: 1 FKPKMR 6

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR68:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	94.3	380	2	C64364
2	33	94.3	412	2	E65146
3	33	94.3	498	2	H71279
4	33	94.3	557	2	B86020
5	33	94.3	623	2	T18250
6	31	88.6	1035	2	S78199
7	30	85.7	229	2	H72667
8	30	85.7	462	2	E65131
9	30	85.7	462	2	B86002
10	30	85.7	606	2	B69805
11	30	85.7	684	2	T31640
12	30	85.7	961	2	S67568
13	30	85.7	962	2	UC5808
14	30	85.7	2204	1	RRNZNV
15	29	82.9	119	2	T36433
16	29	82.9	217	2	H64956
17	29	82.9	232	2	F85810
18	29	82.9	301	2	T24993
19	29	82.9	395	2	H69147
20	29	82.9	412	2	H75484
21	29	82.9	435	2	C86340
22	29	82.9	458	2	H82504
23	29	82.9	462	2	T15052
24	29	82.9	467	2	A59268
25	29	82.9	473	2	G72753
26	29	82.9	477	2	S77373
27	29	82.9	488	2	B82967
28	29	82.9	522	2	T05142
29	29	82.9	554	1	S34607

30	29	82.9	557	2	T39308	asparagine sy.
31	29	82.9	655	2	T30044	hypothetical prote
32	29	82.9	732	2	T50143	pumilio family pro
33	29	82.9	1224	2	S25952	gene cob Inton 3
34	29	82.9	2685	2	T38755	hypothetical prote
35	28	80.0	84	2	T48944	hypothetical prote
36	28	80.0	162	2	S28301	hypothetical prote
37	28	80.0	174	2	D86400	hypothetical prote
38	28	80.0	195	2	T48945	hypothetical prote
39	28	80.0	227	2	B70438	hypothetical prote
40	28	80.0	233	1	VMMJBV	envelope protein E
41	28	80.0	254	1	S10929	transposase - Myco
42	28	80.0	254	2	A49895	transposase - Flav
43	28	80.0	273	2	T37841	probable transloc
44	28	80.0	275	2	JE0183	chitinase (EC 3.2.
45	28	80.0	297	2	E84788	hypothetical prote

ALIGNMENTS

RESULT 1
C64364
Formate hydrogenlyase, subunit 5 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence-revision 13-Sep-1996 #text-change 21-Jul-2000
C:Accession: C64364
R:Butt, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
; Reich, C.J.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
son, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
R:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Moese
R:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannas
A:Reference number: Ab4300; MUID:96337999
A:Accession: C64364
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-380 <RUL>
A:Cross-references: GB:U67501; GB:L77117; MID:92826289; PIDN:ABD95504.1; PID:q159121K
C:Genetics:
A:Map position: REV458767-457625
C:Superfamily: NADH dehydrogenase (ubiquinone) 49k protein

Query Match 94.3% Score 33: DB 2: Length 380:
Best Local Similarity 83.3% Pred. No. 27:
Matches 5: Conservative 0: Mismatches 1: Indels 0: Caps 0:

QY 1 FKPKMR
DB 241 FKPKMR 246

RESULT 2
E65146
hypothetical 47.4 kb protein in rnsB-pit intergenic region - Escherichia coli (strain
H:Alternate names: hypothetical protein f409
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence-revision 17-Sep-1997 #text-change 08-Oct-1999
C:Accession: E65146; S47710
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.
A.: Rose, D.J.; Mau, B.; Shaoh, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: E65146
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-412 <BLAT>
A:Cross-references: GB:AE000425; GB:U00096; MID:92367232; PIDN:AC76515.1; PID:q23672
A:Experimental source: strain K-12, substrain MG1655
R:Plunkett, G.
submitted to the EMBL Data Library, March 1994

A:Reference number: S47666
 A:Accession: S47710
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-398, 'MLSRCAFGST' <PLU>
 A:Cross-references: EMBL:000039; NID:g466582; PIDN:AAB18466.1; PID:g466627
 A:Note: this sequence has been corrected
 C:Genetics:
 A:Gene: yhlL

Query Match 94.3%; Score 33; DB 2; Length 412;
 Best Local Similarity 83.3%; Pred. No. 30;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPMR 6
 DB 168 FKPMR 173

RESULT 3
 H71279
 probable fenn protein - syphilis spirochete
 C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
 C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999
 C:Accession: H71279
 R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
 rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Uterback, T.; McD
 they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
 Science 281, 375-388, 1998
 A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
 A:Reference number: A71250; MUID:98332770
 A:Accession: H71279
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-498 <COL>
 A:Cross-references: GB:AE001251; GB:AE000520; NID:g3323111; PIDN:AAC65773.1; PID:g332311
 A:Experimental source: strain Nichols
 C:Genetics:
 A:Gene: TP0806

Query Match 94.3%; Score 33; DB 2; Length 498;
 Best Local Similarity 83.3%; Pred. No. 36;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPMR 6
 DB 240 FKPMR 245

RESULT 4
 B86020
 hypothetical protein 24888 [imported] - Escherichia coli (strain O157:H7)
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
 C:Accession: B86020
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoultis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:1120551
 A:Accession: B86020
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-557 <STO>
 A:Cross-references: GB:AE005174; NID:g12518181; PIDN:AAG58622.1; GSPDB:GN00145; UMGCP:248
 C:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: 24888

Query Match 94.3%; Score 33; DB 2; Length 557;

Best Local Similarity 83.3%; Pred. No. 40;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 1 FKPMR 6
 DB 189 FKPMR 194

RESULT 5
 T18250
 peptide transport protein - yeast (Candida albicans)
 C:Species: Candida albicans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T18250
 R:Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, November 1998
 A:Reference number: Z18831
 A:Accession: T18250
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-623 <BAR>
 A:Cross-references: EMBL:AL033503; NID:e1341066; PID:e1341077; PIDN:CA22021.1
 C:Genetics:
 A:Gene: plr2
 A:Introns: 475/3

Query Match 94.3%; Score 33; DB 2; Length 623;
 Best Local Similarity 83.3%; Pred. No. 44;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPMR 6
 DB 445 FKPMR 450

RESULT 6
 S78199
 probable maturase protein 3 - fission yeast (Schizosaccharomyces pombe) mitochondrion
 H:Alternate names: gene cob intron 1 protein
 C:Species: mitochondrion Schizosaccharomyces pombe
 C:Date: 29-Jan-1998 #sequence_revision 20-Feb-1998 #text_change 21-Jul-2000
 C:Accession: S78199; S10070; S67378
 R:Lang, B.F.
 submitted to the EMBL Data Library, August 1990
 A:Description: The mitochondrial genome of the fission yeast Schizosaccharomyces pombe
 chizosaccharomyces pombe and Aspergillus nidulans.
 A:Reference number: S78199
 A:Accession: S78199
 A:Molecule type: DNA
 A:Residues: 1-1035 <LAN>
 A:Cross-references: EMBL:X54421
 A:Experimental source: strain ad7-50h
 R:Lang, B.F.; Ahne, F.; Bonen, L.
 J. Mol. Biol. 184, 353-366, 1985
 A:Title: The mitochondrial genome of the fission yeast Schizosaccharomyces pombe. The
 e
 A:Reference number: S07274; MUID:86011547
 A:Accession: S10070
 A:Molecule type: DNA
 A:Residues: 1-1035 <LAN>
 A:Cross-references: EMBL:X02819
 C:Genetics:
 A:Gene: mitochondrion
 A:Genetic code: SGC2
 C:Superfamily: cytochrome b6 homology; cytochrome b homology; plastoquinol--plastoqua
 C:Keywords: heme; iron; metalloprotein; mitochondrion
 F:1-228/region: cytochrome b exon 1 encoded
 F:10-228/region: cytochrome b homology #status atypical <CBH>
 F:10-210/region: cytochrome b6 homology <CB6>
 F:222-228/region: plastoquinol--plastocyanin reductase 17k protein homology #status n
 F:229-1035/region: cytochrome b intron encoded

Query Match 88.6%; Score 31; DB 2; Length 1035;
 Best Local Similarity 66.7%; Pred. No. 1.8e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPMR 6
 :1111
 DB 935 FKPSMK 940

RESULT 7

H72667
 hypothetical protein APE0766 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
 C:Accession: H72667
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix
 A:Reference number: A72450; MUID:99310339

A:Accession: H72667

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-239 <KAM>

A:Cross-references: DDBJ:AP000060; NID:95104188; PIDN:BAA79744.1; PID:95104429

A:Experimental source: strain K1

A:Gene: APE0766

C:Superfamily: Methanobacterium thermoautotrophicum conserved hypothetical protein MTH10

Query Match 85.7%; Score 30; DB 2; Length 239;
 Best Local Similarity 66.7%; Pred. No. 70;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPMR 6
 :1111
 DB 105 FKPSMQ 110

RESULT 8

E65131
 hypothetical 47.5 kD protein in cyst-trps intergenic region - Escherichia coli (strain K1)

C:Species: Escherichia coli
 C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 17-Mar-2000

C:Accession: E65131

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C. A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: E65131

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-462 <BLAT>

A:Cross-references: GB:AE000413; GB:U00096; NID:92367215; PIDN:AC76395.1; PID:92367216;

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: yhfM

C:Superfamily: arginine permease

Query Match 85.7%; Score 30; DB 2; Length 462;
 Best Local Similarity 66.7%; Pred. No. 1.3e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPMR 6
 :1111
 DB 398 YKPLMR 403

RESULT 9

B86002

probable amino acid/amine transport protein yhfM [imported] - *Esche*

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-02

C:Accession: B86002

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, U.; May

iller, L.; Grobeck, E.J.; Davis, N.M.; Lam, A.; Dimalanta, E.; Potamoultis, K.; Apoda

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206531

A:Accession: B86002

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-462 <STNO>

A:Cross-references: GB:AE005174; NID:q12517998; PIDN:AG58478.1; GSPDB:GN00145; UMGCP:

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: yhfM

Query Match 85.7%; Score 30; DB 2; Length 462;
 Best Local Similarity 66.7%; Pred. No. 1.3e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPMR 6
 :1111
 DB 398 YKPLMR 403

RESULT 10

F69805
 conserved hypothetical protein yfix - *Bacillus subtilis*

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999

C:Accession: B69805

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber

C.; Bron, S.; Brouillette, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

A.; Ehrlich, S.D.; Emerson, P.T.; Eutian, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Iulio, M

koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau

Y, M.; Ojawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Porteu

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadleir, Y.; Sato, T.; Scanl

A:Authors: Schleich, S.; Schroeder, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Se

akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*

A:Reference number: A69580; MUID:98044033

A:Accession: B69805

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-606 <KUN>

A:Cross-references: GB:Z99108; GB:AL009126; NID:92633055; PIDN:CA812672.1; PID:ej1828

A:Experimental source: strain 168

C:Genetics:

A:Gene: yfix

Query Match 85.7%; Score 30; DB 2; Length 606;
 Best Local Similarity 66.7%; Pred. No. 1.7e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPMR 6
 :1111
 DB 570 YKPEWR 575

RESULT 11

hypothetical protein Y57A10A.r - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T31640
 R:Smyle, R.
 submitted to the EMBL Data Library, September 1999
 A:Reference number: 221048
 A:Accession: T31640
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-684 <MW>
 A:Cross-references: EMBL:AL117195; NID:e1549729; PIDN:CAB55023.1; CESP:Y57A10A.r
 A:Experimental source: clone Y57A10A
 C:Genetics:
 A:Gene: CESP:Y57A10A.r
 A:introns: 108/3; 132/3; 149/1; 180/3; 323/2; 430/1; 530/1; 584/2; 647/1

Query Match 85.7%; Score 30; DB 2; Length 684;
 Best Local Similarity 66.7%; Pred. No. 1.9e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPMR 6
 1111
 DB 672 FKPMR 677

RESULT 12

S67568
 Probable membrane protein YDL035c - yeast (*Saccharomyces cerevisiae*)
 N:Alternate names: hypothetical protein D2749
 C:Species: *Saccharomyces cerevisiae*
 C:Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 29-Oct-1999
 C:Accession: S67568
 R:Paulin, L.; Saren, A.M.; Laamanen, P.
 submitted to the Protein Sequence Database, July 1996
 A:Reference number: S67560
 A:Accession: S67568
 A:Molecule type: DNA
 A:Residues: 1-961 <PAU>
 A:Cross-references: EMBL:Z74083; NID:g1431014; PID:e252988; PID:g1431015; GSPDB:GN00004;
 A:Experimental source: strain S288C
 C:Genetics:
 A:Gene: MIPS:YDL035c
 A:Map position: 4L
 C:Keywords: transmembrane protein
 F:57-73/Domain: transmembrane #status predicted <TM1>
 F:92-108/Domain: transmembrane #status predicted <TM2>
 F:140-156/Domain: transmembrane #status predicted <TM3>
 F:181-197/Domain: transmembrane #status predicted <TM4>
 F:256-272/Domain: transmembrane #status predicted <TM5>
 F:623-639/Domain: transmembrane #status predicted <TM6>
 F:662-678/Domain: transmembrane #status predicted <TM7>

Query Match 85.7%; Score 30; DB 2; Length 961;
 Best Local Similarity 66.7%; Pred. No. 2.6e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPMR 6
 1111
 DB 156 FKPMR 161

RESULT 13

JC5808
 G:protein-coupled receptor 1 - yeast (*Saccharomyces cerevisiae*)
 C:Species: *Saccharomyces cerevisiae*
 C:Date: 04-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 19-May-2000
 C:Accession: JC5808
 R:Yun, C.W.; Tamaki, H.; Nakayama, R.; Yamamoto, K.; Kumagai, H.
 Biochem. Biophys. Res. Commun. 240, 287-292, 1997
 A:Title: G-protein coupled receptor from yeast *Saccharomyces cerevisiae*.
 A:Reference number: JC5808; MUID:98049822

A:Accession: JC5808
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-962 <YUN>
 C:Comment: This protein monitors the extracellular signal such as nutrition and trans
 C:Genetics:
 A:Gene: gpr1
 F:56-74/Domain: transmembrane #status predicted <TM1>
 F:91-109/Domain: transmembrane #status predicted <TM2>
 F:139-157/Domain: transmembrane #status predicted <TM3>
 F:180-198/Domain: transmembrane #status predicted <TM4>
 F:255-273/Domain: transmembrane #status predicted <TM5>
 F:622-640/Domain: transmembrane #status predicted <TM6>
 F:661-679/Domain: transmembrane #status predicted <TM7>

Query Match 85.7%; Score 30; DB 2; Length 962;
 Best Local Similarity 66.7%; Pred. No. 2.6e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPMR 6
 1111
 DB 156 FKPMR 161

RESULT 14

KRNZNV
 genome polypeptide - Newcastle disease virus (strain Beaudette C)
 N:Contains: RNA-directed RNA polymerase (PC 2.7.7.48)
 C:Species: Newcastle disease virus
 C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 11-Jun-1999
 C:Accession: A26747
 R:Yusoff, K.; Miller, N.S.; Chambers, P.; Emerson, P.T.
 Nucleic Acids Res. 15, 3961-3976, 1987
 A:Title: Nucleotide sequence analysis of the L gene of Newcastle disease virus: homol
 A:Reference number: A93665; MUID:87230982
 A:Accession: A26747
 A:Molecule type: mRNA
 A:Residues: 1-2204 <YUS>
 A:Cross-references: GB:X05399; NID:960937; PIDN:CA28985.1; PID:960939
 C:Genetics:
 A:Gene: L
 C:Superfamily: paraInfluenza virus RNA-directed RNA polymerase
 C:Keywords: ATP; nucleotidyltransferase

Query Match 85.7%; Score 30; DB 1; Length 2204;
 Best Local Similarity 66.7%; Pred. No. 5.8e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPMR 6
 1111
 DB 1844 FKPMR 1849

RESULT 15

T36433
 Hypothetical protein SCF43A.10c - *Streptomyces coelicolor*
 C:Species: *Streptomyces coelicolor*
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: T36433
 R:Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Bartrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, July 1999
 A:Reference number: 221598
 A:Accession: T36433
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-119 <SEE>
 A:Cross-references: EMBL:AL096837; PIDN:CAB48897.1; GSPDB:GN00070; SCOEDB:SCF43A.10c
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCOEDB:SCF43A.10c

Wed Feb 27 12:13:45 2002

us-09-446-109a-11.rpr

Query Match 82.9%; Score 29; DB 2; Length 119;
Best Local Similarity 66.7%; Pred. No. 57;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPPWR 6
1:111
Db 36 FKPPWR 41

Search completed: February 27, 2002, 11:45:20
Job time: 699 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:42:46 : Search time 78.39 Seconds
(without alignments)
2.806 Million cell updates/sec

Title: US-09-446-109A-11

Perfect score: 35
Sequence: 1 FKPKXR 6

Scoring table:

BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 100059 seqs, 3664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	94.3	535	1	YHIL_ECOLI
2	33	94.3	623	1	PTB2_CANAL
3	31	88.6	807	1	YHIL_ECOLI
4	30	85.7	445	1	YHIL_ECOLI
5	30	85.7	2204	1	RRPL_NDVB
6	29	82.9	217	1	YEDK_ECOLI
7	29	82.9	452	1	SPS2_MOUSE
8	29	82.9	477	1	Y264_SYNY3
9	29	82.9	514	1	CSH1_BRAJU
10	29	82.9	522	1	CSH1_ARATH
11	29	82.9	534	1	ESTM_MOUSE
12	28	80.0	151	1	HSPD_BRAJA
13	28	80.0	151	1	HSPH_BRAJA
14	28	80.0	162	1	YLP6_CAREL
15	28	80.0	167	1	CGC6_HUMAN
16	28	80.0	204	1	DSBA_LECPN
17	28	80.0	233	1	VENY_BEV
18	28	80.0	300	1	TYPE_DROME
19	28	80.0	337	1	YHIL_ECOLI
20	28	80.0	340	1	C5AR_GORGO
21	28	80.0	340	1	C5AR_GORGO
22	28	80.0	347	1	C5AR_PANTH
23	28	80.0	350	1	C5AR_MOUSE
24	28	80.0	352	1	C5AR_HUMAN
25	28	80.0	388	1	DXR_BACSU
26	28	80.0	405	1	HS47_CHICK
27	28	80.0	417	1	HS47_HUMAN
28	28	80.0	417	1	HS47_MOUSE
29	28	80.0	417	1	HS47_MOUSE
30	28	80.0	418	1	CBP2_HUMAN
31	28	80.0	469	1	RDXX_RHOSH
32	28	80.0	482	1	C3AR_HUMAN
33	28	80.0	482	1	PUR8_YEAST

Description

P37629 escherichia
P6030 candida alb
P05511 schizosacch
P45539 escherichia
P11205 newcastle d
P67318 escherichia
P97364 mus musculu
P37436 synechocyst
O23738 brassica ju
P46309 arabidopsis
O63880 mus musculu
O69241 bradyrhizob
O86110 bradyrhizob
O03598 caenorhabdi
O93368 homo sapien
P50024 legionella
P27904 berne virus
O93363 drosophila
P37629 escherichia
P79175 gorilla gor
P79240 pan troglod
P30993 mus musculu
P21730 homo sapien
P97520 rattus norv
O31753 bacillus su
P13731 gallus gall
P29043 homo sapien
P19324 mus musculu
P29457 rattus norv
P50454 homo sapien
O01854 rhodobacter
O16581 homo sapien
O05911 saccharomyc

Result No.	Score	Query Match	Length	ID	Description
34	28	80.0	490	1	C883_ARATH
35	28	80.0	501	1	YDF7_SCHPO
36	28	80.0	508	1	GLPK_MYCLE
37	28	80.0	639	1	V70K_PLRVI
38	28	80.0	843	1	POL_MLVAV
39	28	80.0	901	1	PODK_TREPA
40	28	80.0	909	1	VACI_RHISN
41	28	80.0	967	1	V5G0_CLOAB
42	28	80.0	982	1	HBL1_CAREL
43	28	80.0	1133	1	ATX9_TETTH
44	28	80.0	1196	1	POL_MLVAV
45	28	80.0	1275	1	REBC_MYXXA

ALIGNMENTS

RESULT 1
YHIL_ECOLI
STANDARD: PRT: 535 AA.
P37629: P37628: 01-OCT-1994 (Rel. 30, Created)
20-AUG-2001 (Rel. 40, Last sequence update)
20-AUG-2001 (Rel. 40, Last annotation update)
HYPOTHETICAL 61.6 KDA PROTEIN IN RHB-PIT INTERGENIC REGION.
YHIL OR B3489/B3490.
Escherichia coli.
Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
NCBI_TaxID=562;
[1]
SEQUENCE FROM N.A.
STRAIN=K12 / MG1655:
MEDLINE=94316500; PubMed=8041620;
Sofia H.J., Burland V., Daniels D.L., Plunkett G. II, Blattner F.R.,
"Analysis of the Escherichia coli genome. V. DNA sequence of the
region from 76.0 to 81.5 minutes."
Nucleic Acids Res. 22:2576-2586(1994).
[2]
REVIEWS.
STRAIN=K12 / MG1655:
MEDLINE=97426617; PubMed=9278503;
Blattner F.R., Plunkett G. II, Bloch C.A., Perna N.T., Burland V.,
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
Mau B., Shao Y.:
"The complete genome sequence of Escherichia coli K-12."
Science 277:1453-1474(1997).
[3]
CONCEPTUAL TRANSLATION.
Rudd K.E.:
Unpublished observations (JUN-1999).
-1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION. A FRAMESHIFT WAS
INTRODUCED IN POSITION 399 TO PRODUCE THIS ORF.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: U00039; AAB18465.1; ALT_FRAME.
EMBL: U00039; AAB18466.1; ALT_FRAME.
EMBL: AE000425; AAC76514.1; ALT_FRAME.
EMBL: AE000425; AAC76515.1; ALT_FRAME.
DR Ecogen: EGI22227; yhl.
KW Hypothetical protein: Complete proteome.
SEQUENCE 535 AA: 61576 MW: 0046CE53989FDF1 CRC64:

Query Match 94.3% Score 33; DB 1; Length 535;

Best Local Similarity 83.3%; Pred. No. 22;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FRPXR 6
11111
Db 168 FRPXR 173

RESULT 2

PR2_CANAL STANDARD: PRT: 623 AA.

AC P46030: 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DE 01-NOV-1995 (Rel. 32, Last annotation update)

PEPTIDE TRANSPORTER PRT2.

OS Candida albicans (Yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Mitosporic Saccharomycetales; Candida.

OX NCBI_TaxID=5476;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CB5 562 / ATCC 18804;

RA MEDLINE=95291458; PubMed=7773409;

RA Basal M.A., Lubkowitz M.A., Perry J.R., Miller D., Krainer E.,

KA Naider F.R., Becker J.M.;

RT "Cloning of a Candida albicans peptide transport gene."

RL Microbiology 141:1147-1156(1995).

CC - FUNCTION: UPTAKE OF SMALL PEPTIDE.

CC - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

CC - SIMILARITY: BELONGS TO THE PRT2 FAMILY OF TRANSPORTERS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: U09781; AAA0167.1; -

DR InterPro: IPR000109; PRT2.

DR Pfam: PF00854; PRT2.1.

DR PROSITE: PS01022; PRT2.1; 1.

KW Peptide transport; Transport; Transmembrane.

FT TRANSMEM 134 154 POTENTIAL.

FT TRANSMEM 163 183 POTENTIAL.

FT TRANSMEM 191 211 POTENTIAL.

FT TRANSMEM 250 270 POTENTIAL.

FT TRANSMEM 277 297 POTENTIAL.

FT TRANSMEM 385 405 POTENTIAL.

FT TRANSMEM 418 438 POTENTIAL.

FT TRANSMEM 448 468 POTENTIAL.

FT TRANSMEM 499 519 POTENTIAL.

FT TRANSMEM 529 549 POTENTIAL.

FT TRANSMEM 557 577 POTENTIAL.

SQ SEQUENCE 623 AA; 69941 MW; 1601FD3AE21B80EB CRC64;

Query Match 94.3%; Score 33; DB 1; Length 623;

Best Local Similarity 83.3%; Pred. No. 25;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FRPXR 6

11111

Db 445 FRPXR 450

RESULT 3

YH91_SCHIPO STANDARD: PRT: 807 AA.

ID YH91_SCHIPO

AC P45539; P76686;

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE HYPOTHETICAL 91 KDA PROTEIN IN COB INTRON.

OS Schizosaccharomyces pombe (fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OX NCBI_TaxID=4896;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=AD7-50;

RA Lang B.F.;

RL Submitted (AUG-1990) to the EMBL/GenBank/DBJ databases.

CC - MISCELLANEOUS: THIS PROTEIN IS CODED IN THE GROUP-II INTRON OF

CC COB.

CC - SIMILARITY: TO GROUP II INTRON MATURASES.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: X02819; CAA26587.1; -

DR EMBL: X54421; CAA38288.1; -

DR InterPro: IPR003615; HNH.nuc.

DR InterPro: IPR000442; Intron_maturase2.

DR InterPro: IPR00477; RVTse.

DR Pfam: PF01348; Intron_maturase2.1.

DR Pfam: PF00078; rvc.1.

DR SMART: SM00507; HNHc.1.

KW Hypothetical protein; Mitochondrion.

SQ SEQUENCE 807 AA; 91637 MW; AFD70F35E84CA8F CRC64;

Query Match 88.6%; Score 31; DB 1; Length 807;

Best Local Similarity 66.7%; Pred. No. 78;

Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FRPXR 6

11111

Db 707 FRPXR 712

RESULT 4

YHFM_ECOLI STANDARD: PRT: 445 AA.

ID YHFM_ECOLI

AC P45539; P76686;

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE HYPOTHETICAL 47.5 KDA PROTEIN IN CYG-TRPS INTERGENIC REGION.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia

OX NCBI_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-K12 / MG1655:
 RA MEDLINE-9742617; PubMed-9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.:
 RT "The complete genome sequence of *Escherichia coli* K-12."
 RL Science 277:1453-1474(1997).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
 CC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE ARCD/CADR/CERAB/LYSI FAMILY OF
 CC PERMEASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U18997; AAA58167.1; ALT_SEQ.
 DR EMBL: AE000413; AAC76395.1; ALT_INIT.
 DR Ecocore: EG12908; yhfM.
 DR InterPro: IPR002293; AA_rel-permease_1.
 DR InterPro: IPR002027; Amino_acid-permease.
 DR Pfam: PF00324; aa-permeases; 1.
 KM Hypothetical protein; Transport; Transmembrane; Inner membrane;
 KW Complete proteome.
 FT TRANSMEM 10 30 POTENTIAL.
 FT TRANSMEM 38 58 POTENTIAL.
 FT TRANSMEM 93 113 POTENTIAL.
 FT TRANSMEM 121 141 POTENTIAL.
 FT TRANSMEM 155 175 POTENTIAL.
 FT TRANSMEM 181 201 POTENTIAL.
 FT TRANSMEM 236 256 POTENTIAL.
 FT TRANSMEM 273 293 POTENTIAL.
 FT TRANSMEM 334 354 POTENTIAL.
 FT TRANSMEM 355 375 POTENTIAL.
 FT TRANSMEM 389 410 POTENTIAL.
 FT TRANSMEM 417 435 POTENTIAL.
 SO SEQUENCE 445 AA; 47576 MW; B98DF5944A5E1FEF CRC64;

Query Match 85.7%; Score 30; DB 1; Length 445;
 Best Local Similarity 66.7%; Pred. No. 71;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPYWR 6
 DB 381 YKPLMR 386

RESULT 5
 ID REP_L_NDV8 STANDARD: PRT: 2204 AA.
 AC P11205;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE RNA POLYMERASE BETA SUBUNIT (EC 2.7.7.48) (LARGE STRUCTURAL PROTEIN)
 DE (L PROTEIN).
 GN L.
 OS Newcastle disease virus (strain Beaudette C/45) (NDV).
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
 OC Paramyxoviridae; Paramyxovirinae; Rubulaviruses.
 OX NCBI_TaxID=11178;
 RN RP SEQUENCE FROM N.A.
 RX MEDLINE-87230982; PubMed-3035486;
 RA Yusuf K., Millar N.S., Chambers P., Emerson P.T.:
 RT "Nucleotide sequence analysis of the L gene of Newcastle disease
 virus: homologues with Sendai and vesicular stomatitis viruses.";

RL Nucleic Acids Res. 15:3961-3976(1987).
 CC -1- FUNCTION: PROBABLE COMPONENT OF THE ACTIVE POLYMERASE. IT MAY
 CC FUNCTION IN RNA SYNTHESIS, CAPPING, METHYLATION AND POLY(A)
 CC SYNTHESIS OF NEWLY SYNTHESIZED VIRAL MRNAs, RNA EDITING OF THE P
 CC GENE TRANSCRIPT, AND PROTEIN KINASE ACTIVITY.
 CC -1- SIMILARITY: WITH L PROTEIN OF OTHER PARAMYXOVIRUSES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X03399; CAA28985.1; -.
 DR PIR: A26747; RRM2NV.
 DR InterPro: IPR001016; Paramyx_RNA_pol.
 DR Pfam: PF00946; Paramyx_RNA_pol; 1.
 KM Transference: RNA-directed RNA polymerase.
 SO SEQUENCE 2204 AA; 248822 MW; C67B8674D904802C CRC64;

Query Match 85.7%; Score 30; DB 1; Length 2204;
 Best Local Similarity 66.7%; Pred. No. 3e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPYWR 6
 DB 1844 FRPLMR 1849

RESULT 6
 ID YEDK_ECOLI STANDARD: PRT: 217 AA.
 AC P76318;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHEICAL 24.5 KDA PROTEIN IN AMY-A-FLIE INTERGENIC REGION.
 DE YEDK OR B1931.
 GN *Escherichia coli*.
 OS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC *Escherichia*.
 OX NCBI_TaxID=562;
 RN RP SEQUENCE FROM N.A.
 RX MEDLINE-93481452; PubMed-8371104;
 RA Raha M., Kikura M., Kawagishi I., Macnab R.M.:
 RT "Organization of the *Escherichia coli* and *Salmonella typhimurium*
 RT chromosomes between flagellar regions IIIa and IIId, including a
 RT large non-coding region."
 RT J. Gen. Microbiol. 139:1401-1407(1993).
 GN [2]
 SO SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655:
 RX MEDLINE-9742617; PubMed-9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.:
 RT "The complete genome sequence of *Escherichia coli* K-12."
 RL Science 277:1453-1474(1997).
 CC -1- SIMILARITY: TO YEAST YMR114C.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC -----
 DR EMBL: L13279: - NOT_ANNOTATED_CDS.
 DR EMBL: AE000285: AAC7498.1: -
 DR EcoGene: EG13278: yedk.
 DR InterPro: IPR003738: DUF159.
 DR Pfam: PF02586: DUF159.1.
 DR Hypothetical protein: Complete proteome.
 KW CONFLICT 213 TRARS -> NOGAELIOPV (1N REF. 1).
 FT SEQUENCE 217 AA: 24500 MW: 4f26c95db3b02900 CRC64;
 SQ

Query Match 82.9% Score 29: DB 1: Length 217;
 Best Local Similarity 66.7% Pred. NO. 57;
 Matches 4: Conservative 1: Mismatches 1: Indels 0: Gaps 0;

Oy 1 FKPMR 6
 Db 87 FKPLMQ 92

RESULT 7
 SP52_MOUSE STANDARD: PRT: 452 AA.
 AC P97364;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE SELENIDE WATER DIKINASE 2 (EC 2.7.9.3). (SELENOPHOSPHATE SYNTHETASE 2)
 DE (SELENIDIUM DONOR PROTEIN 2).
 GN SP52.
 OS Mus musculus (Mouse).
 OC Eukaryota: Metazoa: Chordata: Vertebrata: Euteleostomi:
 OC Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=96017645; PubMed=7588067;
 RA Guimaraes M.J., Bazan J.F., Zlotnick A., Wiles M.V., Grimaldi J.C.,
 RA Lee F., McClanahan T.;
 RT "A new approach to the study of haematopoietic development in the
 RT yolk sac and embryoid bodies";
 RL Development 121:335-3346(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=97140286; PubMed=8986768;
 RA Guimaraes M.J., Peterson D., Vicari A., Cocks B.G., Copeland N.C.,
 RA Gilbert D.J., Jenkins N.A., Ferrick D.A., Kastelein R., Bazan J.F.,
 RA Zlotnick A.;
 RT "Identification of a novel self homolog from eukaryotes, bacteria,
 RT and archaea: Is there an autoregulatory mechanism in selenocysteine
 RT metabolism?";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:15086-15091(1996);
 CC -1- FUNCTION: SYNTHESIZES SELENOPHOSPHATE FROM SELENIDE AND ATP.
 CC -1- CATALYTIC ACTIVITY: ATP + SELENIDE + H(2)O = AMP + SELENOPHOSPHATE
 CC + PHOSPHATE.
 CC -1- COFACTOR: SELENOCYSTEINE. THE ACTIVE-SITE SELENOCYSTEINE IS
 CC ENCODED BY THE OPAL CODON, UGA.
 CC -1- SIMILARITY: BELONGS TO THE SELENOPHOSPHATE SYNTHETASE 1 FAMILY.
 CC CLASS 1 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U43285: AAC53024.1: -
 CC MCD: MGI:108388: SP52.
 DR InterPro: IPR000728: ATRS-related.
 DR Pfam: PF00586: ATRS: 1.
 DR Transferrase: Selenium: Selenocysteine: ATP-binding.

FT ACT_SITE 63 63 POTENTIAL.
 FT SE_CYS 63 63
 FT SITE 66 66
 FT NP_BIND 322 328 IMPORTANT FOR CATALYTIC ACTIVITY (BY
 FT DOMAIN 2 9 SIMILARITY).
 FT POLY-ALA. 433 440 ATP (POTENTIAL).
 FT POLY-ALA. 433 440
 FT SEQUENCE 452 AA: 47786 MW: 9DA6F7250CFE80E4 CRC64;
 SQ

Query Match 82.9% Score 29: DB 1: Length 452;
 Best Local Similarity 66.7% Pred. NO. 1,1e+02;
 Matches 4: Conservative 0: Mismatches 2: Indels 0: Gaps 0;

Oy 1 FKPMR 6
 Db 48 FSPSMR 53

RESULT 8
 Y264_SYNY3 STANDARD: PRT: 477 AA.
 AC P73436;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL 54.0 KDA PROTEIN SL11464.
 GN SL11464.
 GN Synechocystis sp. (strain PCC 6803).
 OC Bacteria: Cyanobacteria: Chroococcales: Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=97061201; PubMed=8905231;
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hirosewa M., Sugita M., Sasamoto S., Kimura T.,
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naito K.,
 RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
 RA Yamada M., Yasuda M., Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions";
 RL DNA Res. 3:109-136(1996).
 CC -1- SIMILARITY: BELONGS TO THE UPF0061 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D90906: BAA17476.1: -
 DR InterPro: IPR003846: UPF0061.
 DR Pfam: PF02696: UPF0061: 1.
 KW Hypothetical protein: Complete proteome.
 FT SEQUENCE 477 AA: 54041 MW: 81F6899B1A60613C CRC64;
 SQ

Query Match 82.9% Score 29: DB 1: Length 477;
 Best Local Similarity 66.7% Pred. NO. 1.2e+02;
 Matches 4: Conservative 0: Mismatches 2: Indels 0: Gaps 0;

Oy 1 FKPMR 6
 Db 388 FSPSMR 393

RESULT 9
 GSHL_BRAJU STANDARD: PRT: 514 AA.
 ID GSHL_BRAJU
 AC O23736: Q43389;

RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., All J., Berghoff A., Jones K., Drone K., Cotton M., Joshu C.,
 RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
 RA Ma P., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shah R.,
 RA Saby I.R., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
 RA Granat S., Shohdy N., Hasegawa A., Hameed A., Looht M., Johnson A.,
 RA Chen E., Marra M., Mathiesen R., McCombie W.R.,
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 thaliana."
 RT Thaliana."
 RL Nature 402:769-777(1999)
 CC -1- NATURE 402:769-777(1999)
 CC -1- ORTHOPHOSPHATE + GAMMA-L-GLUTAMYL-L-CYSTEINE.
 CC -1- ENZYME REGULATION: FEEDBACK INHIBITION BY GLUTATHIONE.
 CC -1- PATHWAY: FIRST AND RATE-LIMITING STEP IN DE NOVO GLUTATHIONE
 CC BIOSYNTHESIS.
 CC -1- SUBUNIT: MONOMER (PROBABLE).
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST (PROBABLE).
 CC -1- TISSUE SPECIFICITY: ABUNDANT IN LEAVES AND ROOTS.
 CC -1- MISCELLANEOUS: IN PLANTS, THERE ARE TWO FORMS OF THIS ENZYME, ONE
 CC CYTOSOLIC AND THE OTHER CHLOROPLASTIC.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: 229490; CAAB2626.1; -
 DR EMBL: Y09944; CAA71075.1; -
 DR EMBL: AF068299; AAD14544.1; -
 DR EMBL: AL031018; AAD19826.1; -
 DR EMBL: AL161558; CAB79265.1; -
 DR Mendel: 17259; Arabid:2621;17259.
 DR Mendel: 33329; Arabid:2621;33329.
 DR Mendel: 36232; Arabid:2621;36232.
 KW Glutathione biosynthesis; Ligation; Chloroplast; Transic peptide.
 FT TRANSIT 1 ?
 FT CHAIN 1 ?
 FT VARIANT 22 22
 FT VARIANT 48 48
 FT CONFLICT 491 522
 FT VRTGVTPEAKLEWYNGECOSVDPVEELY -> CONRS
 FT YACCEALGDYQMRMGTRKRSRVRRAVALRKWDVNRCL
 FT (IN REF. 1).
 SO SEQUENCE 522 AA: 58562 MW: CCBP13C6F44E0EF7 CMC64;

Query Match 82.9%; Score 29; DB 1; Length 522;
 Best Local Similarity 66.7%; Pred. No. 1.3e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Caps 0;

OY 1 FKPKWR 6
 DB 212 FQPKWR 217

RESULT 11
 ESTM_MOUSE STANDARD: PRT; 554 AA.
 AC 063880:
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE LIVER CARBOXYL ESTERASE PRECURSOR (EC 3.1.1.1) (ES-MALE) (ESTERASE-31).
 GN ES31.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN 111
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=BAUB/CJ X DBA/2J; TISSUE=Liver;

EX MEDLINE=93326638; PubMed=7916639;
 RA Alda K., Moore R., Negishi M.;
 RT "Cloning and nucleotide sequence of a novel, male-predominant
 RT Carboxylesterase in mouse liver."
 RL Biochim. Biophys. Acta 1174:72-74(1993).
 CC -1- FUNCTION: INVOLVED IN THE DETOXIFICATION OF XENOBIOTICS AND IN
 CC THE ACTIVATION OF ESTER AND AMIDE PRODRUGS.
 CC -1- CATALYTIC ACTIVITY: A CARBOXYLIC ESTER + H(2)O = AN ALCOHOL
 CC + A CARBOXYLIC ANION.
 CC -1- SUBCELLULAR LOCATION: MICROSOMAL MEMBRANE, LUMEN OF ENDOPLASMIC
 CC RETICULUM.
 CC -1- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: S64130; AAB27606.1; -
 DR HSSP: P21836; 1MAH.
 DR MKD: MGT102773; ES31.
 DR InterPro: IPR002018; Carboxylesterase-B.
 DR InterPro: IPR000379; Est_1ip.thioest_actsite.
 DR Pfam: PF00135; Coesterase; 1.
 DR PROSITE: PS00122; CARBOXYLESTERASE_B_1; 1.
 DR PROSITE: PS00941; CARBOXYLESTERASE_B_2; 1.
 KW Glycoprotein; Hydrolase; Serine esterase; Endoplasmic reticulum;
 KW Signal; Multigene family.
 FT SIGNAL 14
 FT CHAIN 15 554
 FT ACT_SITE 215 215
 FT ACT_SITE 443 443
 FT DISULFID 83 110
 FT DISULFID 267 278
 FT SITE 551 554
 FT CARBOHYD 294 304
 FT CARBOHYD 300 300
 FT CARBOHYD 381 381
 FT SEQUENCE 554 AA: 61509 MW: 028D89B8B06F3CANAD CRC64;

Query Match 82.9%; Score 29; DB 1; Length 554;
 Best Local Similarity 80.0%; Pred. No. 1.3e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FKPKW 5
 DB 434 FKPAW 438

RESULT 12
 HSPD_BRAJA STANDARD: PRT; 151 AA.
 AC 069241:
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE SMALL HEAT SHOCK PROTEIN HSPD.
 GN HSPD.
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 CC Bradyrhizobium group; Bradyrhizobium.
 OX NCBI_TaxID=375;
 RN 111
 RP SEQUENCE FROM N.A.
 RA Narberhaus F., Weiglhofer W., Fischer H.M., Hennecke H.;
 RT "Identification of the Bradyrhizobium japonicum desr gene as part of
 RT an operon containing small heat shock protein genes."
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)

```

CC      FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: AJ003064; CAA05835.1; -
DR      InterPro: IPR002068; CrystalIn_HSP20.
DR      Pfam: PF00011; HSP20; 1.
DR      PROSITE: PS01031; HSP20; 1.
KM      Heat shock; Multigene family.
SQ      SEQUENCE 151 AA; 17272 MW; FB44EF94FB599EE4 CRC64;

Query Match
Best Local Similarity 80.0%; Score 28; DB 1; Length 151;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      1 FKPYMR 6
        1 1 1
        6 FSPLMR 11

Db
RESULT 13
HSPH_BRAJA STANDARD; PRT; 151 AA.
AC 086110;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE SMALL HEAT SHOCK PROTEIN HSPH.
OS HSPH.
OC Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bradyrhizobium group; Bradyrhizobium.
OX NCBI_TaxID=375;
RN 11
RP SEQUENCE FROM N.A.
RA Muenchbach M., Nocker A., Narberhaus F.;
RT "Occurrence of a superfamily of small heat shock proteins in
RT Bradyrhizobium japonicum and other Rhizobium species: a plant-like
RT phenomenon.";
RL Submitted (Aug-1998) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
CC FAMILY.
CC -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: AJ010144; CAA09014.1; -
DR      InterPro: IPR002068; CrystalIn_HSP20.
DR      Pfam: PF00011; HSP20; 1.
DR      PROSITE: PS01031; HSP20; 1.
KM      Heat shock; Multigene family.
SQ      SEQUENCE 151 AA; 17098 MW; DFE9809310AF6A28 CRC64;

Query Match
Best Local Similarity 80.0%; Score 28; DB 1; Length 151;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      1 FKPYMR 6
        1 1 1
        6 FSPLMR 11

Db

```

```

CC      RESULT 14
CC      YLF6_CAEBL STANDARD; PRT; 162 AA.
CC      ID YLF6_CAEBL
CC      AC 003586;
CC      DT 01-FEB-1994 (Rel. 28, Created)
CC      DT 01-FEB-1994 (Rel. 28, Last sequence update)
CC      DT 30-MAY-2000 (Rel. 39, Last annotation update)
CC      DE HYPOTHEICAL 18.5 KDA PROTEIN C40H1.6 IN CHROMOSOME III.
CC      CN C40H1.6.
CC      OS Caenorhabditis elegans.
CC      CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
CC      CC Rhabditidae; Pelodierinae; Caenorhabditis.
CC      OX NCBI_TaxID=6239;
CC      RN 11
CC      RP SEQUENCE FROM N.A.
CC      RC STRAIN-BRISTOL N2;
CC      RX MEDLINE=94150718; PubMed=7906398;
CC      RA Wilsson R., Ainscough R., Anderson K., Baynes C., Berks M.,
CC      RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
CC      RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,
CC      RA Johnston L., Jones M., Kershaw J., Lloyd C., Mortimore B., O'Callaghan M.,
CC      RA Paterson J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
CC      RA Sims M., Smailton N., Smith A., Smith M., Sonnenhammer E., Staden R.,
CC      RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
CC      RA Waterson R., Watson A., Weltsch L., Wilkinson-Sproat J.,
CC      RA Wohldman P.;
CC      RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
CC      RT elegans.";
CC      RL Nature 368:32-38(1994).
CC      CC -1- SIMILARITY: STRONG, TO HUMAN CGI-126.
CC -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: Z19154; CAA79557.1; -
CC      PIR: S28301; S28301.
CC      DR WormRep: C40H1.6; CE00114.
CC      KW Hypothetical protein.
CC      SQ SEQUENCE 162 AA; 18537 MW; 8DC03CE38CE79D55 CRC64;

Query Match
Best Local Similarity 80.0%; Score 28; DB 1; Length 162;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1 FKPYMR 5
        1 1 1
        120 FKPLM 124

Db
RESULT 15
CCCG_HUMAN STANDARD; PRT; 167 AA.
ID CC66_HUMAN
AC 09Y3C8;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHEICAL PROTEIN CGI-126.
OS Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CC OX NCBI_TaxID=9606;
CC RN 11
CC RP SEQUENCE FROM N.A.

```

RX MEDLINE=20272150: PubMed=10810093:
RA Lal C.-H., Chou C.-Y., Chang L.-Y., Liu C.-S., Lin W.-C.:
RT "Identification of novel human genes evolutionarily conserved in
RL Caenorhabditis elegans by comparative proteomics."
Genome Res. 10:703-713(2000).
CC -1- SIMILARITY: STRONG, TO C.ELEGANS C40H1.6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AF151884; AAD34121.1; -.
DR Hypothetical protein.
KW
SQ SEQUENCE 167 AA: 19458 MW: 1675D9187DC43E14 CRC64;

Query Match 80.0%; Score 28; DB 1; Length 167;
Best Local Similarity 80.0%; Pred. No. 70;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 FRPXM 5
DB 121 FRPLW 125

Search completed: February 27, 2002, 11:42:46
Job time: 545 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:50:11 : Search time 281.76 seconds
(Without alignments)
3.115 Million cell updates/sec

Title: US-09-446-109a-11
Perfect score: 35
Sequence: 1 FKPMR 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP_archaea:*
2: SP_bacteria:*
3: SP_fungi:*
4: SP_human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhc:*
8: SP_organelle:*
9: SP_phage:*
10: SP_plant:*
11: SP_rodent:*
12: SP_virus:*
13: SP_vertebrate:*
14: SP_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	34	97.1	250	2	09EUM2	09EUM2 corynebacte
2	33	94.3	380	1	057935	057935 methanococ
3	33	94.3	498	2	083784	083784 treponema p
4	33	94.3	623	3	09URL7	09UR17 candida alb
5	33	94.3	917	5	09VYP8	09VYP8 drosophila
6	33	94.3	917	5	09VYP4	09VYP4 oryza sativ
7	31	88.6	221	10	09XJ12	09XJ12 mus musculu
8	30	85.7	204	11	09D3F2	09D3F2 mus musculu
9	30	85.7	239	1	09YED3	09YED3 aeropyrum p
10	30	85.7	440	11	09DPR1	09DPR1 mus musculu
11	30	85.7	606	2	031566	031566 bacillus su
12	30	85.7	610	2	052961	052961 bacillus su
13	30	85.7	684	5	09NA80	09NA80 caenorhabd
14	30	85.7	961	3	012361	012361 saccharomyc
15	30	85.7	1232	10	09LJQ1	09LJQ1 arabidopsis
16	30	85.7	2204	12	090341	090341 newcastlle d
17	30	85.7	2204	12	09NMH6	09NMH6 newcastlle d
18	30	85.7	2204	12	09DL03	09DL03 newcastlle d
19	30	85.7	2454	3	09UVP2	09UVP2 emericella

20	30	85.7	2454	3	09UVP6	09UVP6 emericella
21	29	82.9	119	2	09XAA5	09XAA5 streptomyc
22	29	82.9	138	2	09ZEM4	09ZEM4 klebsiella
23	29	82.9	143	5	096402	096402 plasmidium
24	29	82.9	162	2	09CJ00	09CJ00 pasteurella
25	29	82.9	164	5	09NJV1	09NJV1 giardia lam
26	29	82.9	224	5	09NZY6	09NZY6 caenorhabd
27	29	82.9	233	2	09FZR9	09FZR9 streptomyc
28	29	82.9	244	10	09RS78	09RS78 triticum ae
29	29	82.9	262	5	09NND2	09NND2 leishmania
30	29	82.9	286	2	033353	033353 mycobacteri
31	29	82.9	301	5	022568	022568 caenorhabd
32	29	82.9	395	1	026470	026470 methanobact
33	29	82.9	398	2	09A710	09A710 caulobacter
34	29	82.9	412	2	09RME3	09RME3 delnoccocc
35	29	82.9	436	10	09LH75	09LH75 arabidopsis
36	29	82.9	436	10	082665	082665 brassica ju
37	29	82.9	445	3	09C1B2	09C1B2 gibberella
38	29	82.9	447	3	094088	094088 fusarium sp
39	29	82.9	447	3	09C1B9	09C1B9 fusarium sp
40	29	82.9	458	2	09KMB9	09KMB9 vibrio chol
41	29	82.9	462	10	040414	040414 nicotiana s
42	29	82.9	467	2	052729	052729 rhizobium e
43	29	82.9	473	1	09YGB1	09YGB1 aeropyrum p
44	29	82.9	488	2	09HFC7	09HFC7 pseudomonas
45	29	82.9	557	3	042902	042902 schizosacch

ALIGNMENTS

RESULT 1
ID 09EUM2 PRELIMINARY: PRT: 250 AA.
AC 09EUM2:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE YGCAL.
GN YGCAL.
US Corynebacterium glutamicum (Brevibacterium flavum).
OC Plasmid R-plasmid PCG4.
OC Bacteria: Firmicutes: Actinobacteria: Actinobacteridae:
OC Actinomycetales: Corynebacterineae: Corynebacteriaceae:
OC Corynebacterium.
OX NCBI_TaxID=1718:
RN [1]
RP SEQUENCE FROM N.A.
PC STRAIN-ATCC31830:
KA Trauch A., Puenher A., Kalinowski J.:
RT "DNA sequence and genetic organization of the integron-carrying R-
RT plasmid PCG4 of Corynebacterium glutamicum.";
RL Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.
NR EMBL: AF164956; AAC00294.1;
DR InterPro: IPR002145; CopG_HTH_4...
DR Pfam: PF01402; HTH_4; 1.
KW Plasmid.
SQ SEQUENCE 250 AA: 28804 MW: 195C86367B483157 CRC64:

Query Match 97.1%: Score 34: DB 2: Length 250:
Best Local Similarity 83.3%: Pred. No. 35:
Matches 5: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

OY 1 FKPMR 6
ID 214 FKPMR 219

RESULT 2
ID 057935 PRELIMINARY: PRT: 380 AA.
AC 057935:

```

DT 01-JAN-1998 (TEMBLrel. 05, Created)
DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE PUTATIVE FORMATE HYDROGENLYASE SUBUNIT 5 (FHL SUBUNIT 5) (HYDROGENASE-
DE 3 COMPONENT E).
OS Methanococcus jannaschii.
GN M00515.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190.
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.C., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.T.,
RA Overbeek R., Kirkness E.F., Weissink J.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhlman J.L., Nguyen D.,
RA Uitterlinden T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA *Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.
RT Science 273:1058-1073(1996).
CC -1- COFACTOR: NICKEL (BY SIMILARITY).
CC -1- PATHWAY: HYDROGEN METABOLISM: FHL PATHWAY (BY SIMILARITY).
CC -1- SUBUNIT: FHL COMPLEXES OF A FORMATE DEHYDROGENASE, UNIDENTIFIED
CC ELECTRON CARRIERS AND A HYDROGENASE (ISOENZYME 3). IN THIS NON-
CC ENERGY CONSERVING PATHWAY MOLECULAR HYDROGEN AND CARBODIOXIDE FROM
CC FORMATE ARE RELEASED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE COMPLEX I 49 KDA SUBUNIT FAMILY.
DR FMBU: U67501; AAB98504.1;
DR TIGR: M00515;
DR InterPro: IPR001135; Complex_49kd.
DR InterPro: IPR001501; Nifese_Hases.
DR Pfam: PF00346; Complex_49kd; 1.
DR Pfam: PF00374; Nifese_Hases; 1.
DR PROSITE: PS00535; COMPLEX_49K; UNKNOWN_1.
DR PROSITE: PS00507; NICKENASE_L_1; UNKNOWN_1.
DR Hypothetical protein: Oxidoreductase; NAD; Iron-sulfur; 4Fe-4S;
KW Nickel; Complete proteome.
SQ SEQUENCE 380 AA; 43638 MW; 3EADI6D67834A76F CRC64;

Query Match 94.3% Score 33; DB 1; Length 380;
Best Local Similarity 83.3% Pred. No. 81;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPYMR 6
DB 241 FKPYMR 246

RESULT 3
ID 083784 PRELIMINARY; PRT; 498 AA.
AC 083784;
DT 01-NOV-1998 (TEMBLrel. 08, Created)
DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE FEMA PROTEIN, PUTATIVE.
DE TP0806.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=98337770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.C.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,

```

```

RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Uitterlinden T.,
RA McDonald L., Attlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RA *Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.
RT Science 281:375-388(1998).
RL EMBL: AE001251; AAC65773.1;
DR TIGR: TP0806;
DR InterPro: IPR003447; FemAB.
DR Pfam: PF02388; FemAB; 1.
KW Complete proteome.
SQ SEQUENCE 498 AA; 56074 MW; 81710E41264431B8 CRC64;

Query Match 94.3% Score 33; DB 2; Length 498;
Best Local Similarity 83.3% Pred. No. 1e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPYMR 6
DB 240 FKPYMR 245

RESULT 4
ID 09URL7 PRELIMINARY; PRT; 623 AA.
AC 09URL7;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE PEPTIDE TRANSPORT PROTEIN.
DE PTR2.
GN Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN=1161;
RA Murphy L., Harris D.;
RA Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RL 121
RN 121
RC SEQUENCE FROM N.A.
RC STRAIN=1161;
RA Bartell B.G., Rajandream M.A.;
RA Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN 131
RP SEQUENCE FROM N.A.
RC STRAIN=1161;
RX MEDLINE=97435544; PubMed=9290243;
RA Tait E., Simon M.C., King S., Brown A.J., Gow N.A.R., Shaw D.J.;
RA "A Candida albicans Genome Project: Cosmid Contigs, Physical Mapping,
RT and Gene Isolation."
RT Fungal Genet. Biol. 21:308-314(1997).
RL EMBL: AL033503; CA22021.1;
DR InterPro: IPR000109; PTR2.
DR Pfam: PF00854; PTR2; 1.
DR PROSITE: PS01022; PTR2_1; 1.
DR PROSITE: PS01023; PTR2_2; UNKNOWN_1.
SQ SEQUENCE 623 AA; 69943 MW; 8543A3B1F7E7363E CRC64;

Query Match 94.3% Score 33; DB 3; Length 623;
Best Local Similarity 83.3% Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPYMR 6
DB 445 FKPYMR 450

RESULT 5

```

09XYP8
ID 09XYP8 PRELIMINARY: PRT: 917 AA.
AC 09XYP8:
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GAMMA-TUBULIN RING PROTEIN DGRIP91.
GN L(1)DD4 OR DGRIP91 OR CG10988.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=99156983; PubMed=10037793;
RA Oegema K., Miese C., Martin O., Milligan R.A., Iwanatsu A.,
RA Mitchelson T.J., Zheng Y.,
RT "Characterization of two related Drosophila gamma-tubulin complexes
RT that differ in their ability to nucleate microtubules.";
RL J. Cell Biol. 144:721-733(1999).
DR EMBL: AF118380; AAD27817.1; -
DR FlyBase: Fgn0001612; 1(1)dd4.
DR InterPro: IPR000634; dehydratase_ser_thr.
DR PROSITE: PS00165; DEHYDRATASE_SER_THR. 1.
SQ SEQUENCE 917 AA: 103819 MW: FBR2FA1E35049D27 CRC64;

Query Match 94.3%; Score 33; DB 5; Length 917;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FRPXR 6
Db 657 FRPXR 662

RESULT 6
09Y94 PRELIMINARY: PRT: 917 AA.
AC 09Y94:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CRIP1 PROTEIN.
GN L(1)DD4 OR CG10988.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer J.R., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolishakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burdits K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Goddek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mallet B., McIntosh N.V., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleib J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Vitskas R., Teclor C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003493; AAF48309.1; -
DR FlyBase: Fgn0001612; 1(1)dd4.
DR InterPro: IPR000634; dehydratase_ser_thr.
DR PROSITE: PS00165; DEHYDRATASE_SER_THR. 1.
SQ SEQUENCE 917 AA: 103706 MW: 6AE88C211D2568B CRC64;

Query Match 94.3%; Score 33; DB 5; Length 917;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FRPXR 6
Db 657 FRPXR 662

RESULT 7
09X12 PRELIMINARY: PRT: 221 AA.
AC 09X12:
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE HYPOTHEICAL PROTEIN.
GN Oryza sativa (Rice).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Erythroidae; Oryzae; Oryza.
OX NCBI_TaxID=4570;
RN 11
RP SEQUENCE FROM N.A.
RX STRAIN=CV. NIPPONBARE;
RA Sasaki T., Nagamura Y., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC
RT clone:POB-114.1.";
FL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB023482; BAA70741.1; -
SQ SEQUENCE 221 AA: 24502 MW: DC24FF370660BC15 CRC64;

Query Match 88.6%; Score 31; DB 10; Length 221;
Best Local Similarity 66.7%; Pred. No. 1.2e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FRPXR 6
Db 203 FRPXR 208

RESULT 8
09D3F2 PRELIMINARY: PRT: 204 AA.
AC 09D3F2:

DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE 5830420C20R1K PROTEIN.
 CN 5830420C20R1K.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=THYMUS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Mashio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Batsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitlaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL: AK017942; BMB31014.1;
 DR MGD: 1923275; 5830420C20R1K
 SQ SEQUENCE 204 AA; 22892 MW; 98AF03DF7C7B066 CRC64;

Query Match 85.7%; Score 30; DB 11; Length 204;
 Best Local Similarity 66.7%; Pred. No. 1.7e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FRPXR 6
 Db 29 FRPXR 34

RESULT 9
 Q9YE03 PRELIMINARY; PRT: 239 AA.
 AC Q9YE03;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE HYPOHETICAL 24.8 KDA PROTEIN APE0766.
 GN APE0766.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
 OC Aeropyrum.
 NCBI_TaxID=56636;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KL;
 RX MEDLINE=99310339; PubMed=10382966;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Itakawa Y.,
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 RA Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 crenarchaeon, Aeropyrum pernix KL.";
 RL DNA Res. 6:83-101(1999).

DR EMBL: AP000060; BAA79744.1;
 DR InterPro: IPR001454; Hydrolase.
 DR InterPro: IPR000150; Hypothet_cof.
 DR Pfam: PF00702; Hydrolase; 1.
 DR PROSITE: PS01229; COF_2; UNKNOWN_1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 239 AA; 24758 MW; EA30A3CDF86B3DE CRC64;

Query Match 85.7%; Score 30; DB 1; Length 239;
 Best Local Similarity 66.7%; Pred. No. 1.9e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FRPXR 6
 Db 105 FRPXR 110

RESULT 10
 Q9D9R1 PRELIMINARY; PRT: 440 AA.
 AC Q9D9R1;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE ADULT MALE TESTIS CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,
 DE CLONE:1700030N20, FULL INSERT SEQUENCE.
 CN 5830420C20R1K.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Mashio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Batsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitlaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL: AK006565; BMB24655.1;
 DR MGD: 1923275; 5830420C20R1K.
 SQ SEQUENCE 440 AA; 49566 MW; 5B5DA87E19879644 CRC64;

Query Match 85.7%; Score 30; DB 11; Length 440;
 Best Local Similarity 66.7%; Pred. No. 3.5e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FRPXR 6
 Db 39 FRPXR 44

RESULT 11
 Q31566 PRELIMINARY; PRT: 606 AA.
 ID Q31566


```

AC 031566;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE YF1X PROTEIN.
GN Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RC MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borstis R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Ertlan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,
RA Gilm S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koeter P., Konigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mello R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudaga B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scottone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenberg M., Vannier F., Vassarotti A.,
RA Viarl A., Wambutt R., Wedler E., Wedler H., Weitzenger T.,
RA Winters P., Wipet A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RC Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RA Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z99108; CAB12672.1; -
KW Complete proteome.
SQ SEQUENCE 606 AA: 68677 MW: 4858058922ED0CA CRC64;

```

Query Match 85.7%; Score 30; DB 2; Length 606;
Best Local Similarity 66.7%; Pred. No. 4.7e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

Qy 1 FKPKWR 6
Db 570 YKPKWR 575

```

RESULT 12
ID 052961 PRELIMINARY; PRT; 610 AA.
AC 052961;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE YF1X.
OC Bacillus subtilis.

```

OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97101647; PubMed=8946165;
RA Yamamoto H., Uchiyama S., Sekiguchi J.;
RT "Cloning and sequencing of a 27.8-kb nucleotide sequence of the 79
RT degrees-81 degrees region of the Bacillus subtilis genome containing
RT the sspe locus."
RL DNA Res. 3:257-262(1996).
DR EMBL: D85082; BAA24464.1; -
SQ SEQUENCE 610 AA: 69089 MW: 34094DD5CA72FF66 CRC64;

```

Query Match 85.7%; Score 30; DB 2; Length 610;
Best Local Similarity 66.7%; Pred. No. 4.7e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

Qy 1 FKPKWR 6
Db 574 YKPKWR 579

```

RESULT 13
ID 09NA80 PRELIMINARY; PRT; 684 AA.
AC 09NA80;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE Y57A10A.20 PROTEIN.
GN Y57A10A.20.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Rhabditida; Rhabditidae;
OC Rhabditidae; Meloidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Smye R.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
KA none.
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology."
EL Science 282:2012-2018(1998).
DR EMBL: AL117195; CAB55023.1; -
SQ SEQUENCE 684 AA: 79659 MW: F0FAFAR28EAF687 CRC64;

Query Match 85.7%; Score 30; DB 5; Length 684;
Best Local Similarity 66.7%; Pred. No. 5.3e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

Qy 1 FKPKWR 6
Db 672 FKPKWR 677

```

RESULT 14
ID 012361 PRELIMINARY; PRT; 961 AA.
AC 012361;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HYPOTHETICAL 110.7 KDA PROTEIN YD1035C.
GN GPRI OR D2749 OR YD1035C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OX NCBJ_TaxID=4932;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN-ALPHA 528C;
 RA Paulin L., Saren A.M.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 RN 12
 RP SEQUENCE FROM N.A.
 RA Paulin L., Saren A.M., Laamanen P.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN 13
 RP SEQUENCE FROM N.A.
 RA MIPS;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: 271781; CAA96454.1; -;
 DR EMBL: 274083; CAA98593.1; -;
 DR SCD: S0002193; GPRI.
 KW Hypothetical protein.
 SQ SEQUENCE 961 AA; 110708 MW; 9889D857872A4209 CRC64;

Query Match 85.7%; Score 30; DB 3; Length 961;
 Best Local Similarity 66.7%; Pred. NO. 7.3e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Caps 0;

OY 1 FKPMX 6
 1111
 DB 156 FKPMX 161

RESULT 15
 09LJ01 PRELIMINARY; PRT: 1232 AA.
 AC 09LJ01;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE G81AAD55299.1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBJ_TaxID=3702;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RA Kaneo T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RN 12
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RX PubMed-10907853;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the regions of 4,251,695 bp covered by ninety PL,
 RT TAC and BAC clones."
 RL DNA Res. 7:217-221(2000).
 CC 1- SIMILARITY: CONTAINS A RING-TYPE ZINC FINGER.
 DR EMBL: AP000414; BAB01179.1; -;
 DR InterPro: IPR000194; ATPase_alpha_beta.
 DR InterPro: IPR000345; CYC_heme_bind.
 DR InterPro: IPR001841; Znf_fing.
 DR Pfam: PF00097; zf-C3HC4; 1.
 DR SMART: SM00184; RING; 1.
 DR PROSITE: PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
 DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
 KW Zinc-finger.
 SQ SEQUENCE 1232 AA; 139223 MW; D3770C4A9D699207 CRC64;

Query Match 85.7%; Score 30; DB 10; Length 1232;
 Best Local Similarity 66.7%; Pred. NO. 9.2e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Caps 0;

CY 1 FKPMX 6
 1111
 DB 902 FKPMX 907

Search completed: February 27, 2002, 11:50:11
 Job time: 990 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:36:05 : Search time 132.19 Seconds
(without alignments)
1.021 Million cell updates/sec

Title: US-09-446-109A-11

Perfect score: 35

Sequence: 1 FKPMR 6

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, AA: *
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep: *
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/PTUS.COMB.pep: *
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	94.3	6	1	US-08-215-137-13
2	28	80.0	13	4	US-08-915-314-26
3	28	80.0	16	1	US-08-079-051-2
4	28	80.0	16	5	PCT-US94-06994-2
5	28	80.0	115	3	US-08-513-974B-351
6	28	80.0	243	2	US-08-829-110-3
7	28	80.0	304	1	US-08-118-270-35
8	28	80.0	304	5	PCT-US93-08528-35
9	28	80.0	350	2	US-08-458-970A-9
10	28	80.0	463	3	US-09-082-310-1
11	28	80.0	482	2	US-08-876-874-2
12	28	80.0	496	3	US-08-906-769-113
13	28	80.0	496	3	US-08-906-616-113
14	28	80.0	496	4	US-08-817-795-113
15	28	80.0	496	4	US-08-639-075A-113
16	28	80.0	496	4	US-09-012-431-113
17	28	80.0	496	4	US-09-012-692-113
18	28	80.0	496	4	US-08-806-613-113
19	28	80.0	496	5	PCT-US93-14442A-113
20	28	80.0	592	3	US-08-991-813-2
21	27	77.1	157	4	US-09-461-474-14
22	27	77.1	355	2	US-08-666-367B-6
23	27	77.1	355	4	US-08-143-438-6
24	27	77.1	448	4	US-09-461-474-8
25	27	77.1	474	4	US-09-461-474-10
26	27	77.1	527	3	US-08-907-229-2
27	27	77.1	566	2	US-08-666-367B-5

28	27	77.1	566	4	US-09-143-438-5	Sequence 5, Appl
29	27	77.1	1040	2	US-08-254-989-2	Sequence 2, Appl
30	27	77.1	1536	4	US-09-413-814-10	Sequence 10, Appl
31	27	77.1	2232	4	US-09-091-219-25	Sequence 25, Appl
32	27	77.1	2247	4	US-09-091-219-2	Sequence 2, Appl
33	26	74.3	10	2	US-08-704-655-20	Sequence 20, Appl
34	26	74.3	37	1	US-08-6,5,5,43B-4	Sequence 4, Appl
35	26	74.3	42	2	US-08-766-858A-27	Sequence 27, Appl
36	26	74.3	209	4	US-09-164-193-8	Sequence 8, Appl
37	26	74.3	273	2	US-08-997-080-75	Sequence 75, Appl
38	26	74.3	273	2	US-08-997-362-75	Sequence 75, Appl
39	26	74.3	273	3	US-08-873-970-75	Sequence 75, Appl
40	26	74.3	273	4	US-09-095-855-75	Sequence 75, Appl
41	26	74.3	336	4	US-09-334-601-13	Sequence 13, Appl
42	26	74.3	370	2	US-08-997-080-194	Sequence 194, App
43	26	74.3	370	2	US-08-997-362-194	Sequence 194, App
44	26	74.3	370	4	US-09-095-855-194	Sequence 194, App
45	26	74.3	417	1	US-08-351-981-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-08-215-137-13
Sequence 13, Application US/08215137
Patent No. 5614370
GENERAL INFORMATION:
APPLICANT: Konteatis, Zenon
APPLICANT: Siciliano, Salvatore J
APPLICANT: Spinger, Martin S
TITLE OF INVENTION: ASSAY TO IDENTIFY HUMAN C5A ANTAGONISTS
TITLE OF INVENTION: AND AGONISTS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: 126 Lincoln Ave., P.O. Box 2000
CITY: Rahway
STATE: New Jersey
COUNTRY: United States of America
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/215,137
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: P. J. G. Gerard H.
REGISTERED NUMBER: 35,746
REFERENCE: KEY NUMBER: 19108
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-3901
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ. ID NO. 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 6 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /label= BH
OTHER INFORMATION: /note= either the natural phenylalanine amino
OTHER INFORMATION: terminus of the Bolton-Hunter modified peptide

OTHER INFORMATION: having the group 3-(p-hydroxyphenyl)propionyl group
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /label=dcha
OTHER INFORMATION: /note="d-cyclohexylalanine"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 6
OTHER INFORMATION: /label=darg
OTHER INFORMATION: /note="d-arginine"
US-08-215-137-13

Query Match 94.3%; Score 33; DB 1; Length 6;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FKPMR 6
11111
DB 1 FKPMR 6

RESULT 2
US-08-915-314-26
Sequence 26, Application US/08915314
Patent No. 6180604
GENERAL INFORMATION:
APPLICANT: Fraser, Janet R.
APPLICANT: West, Michael H.P.
APPLICANT: Krieger, Timothy J.
APPLICANT: Taylor, Robert
APPLICANT: Erile, Douglas
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
NUMBER OF SEQUENCES: 90
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,314
FILING DATE: 20-AUG-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: NO. 6180604 Leburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 660081.405
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: Linear
US-08-915-314-26

Query Match 80.0%; Score 28; DB 4; Length 13;
Best Local Similarity 66.7%; Pred. No. 8.3;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 1 FKPMR 6

DB 4 WKPMR 9
11111

RESULT 3
US-08-079-051-2
Sequence 2, Application US/08079051
Patent No. 5480974
GENERAL INFORMATION:
APPLICANT: MORGAN, EDWARD L.
APPLICANT: EMBER, JULIA A.
APPLICANT: HUGLI, TONY E.
TITLE OF INVENTION: ANTIBODIES TO HUMAN C5a RECEPTOR
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Spensley Horn Juba & Lubitz
STREET: 1880 Century Park East, Suite 500
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/079,051
FILING DATE: 18-JUN-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Wetherell, Jr., Ph.D., John W.
REGISTRATION NUMBER: 31,678
REFERENCE/DOCKET NUMBER: PD-2790
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
CLONE: C5aR(13-148)
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..16
US-08-079-051-2

Query Match 80.0%; Score 28; DB 1; Length 16;
Best Local Similarity 80.0%; Pred. No. 10;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPMR 5
11111
DB 7 FKPMR 11

RESULT 4
PCT-US94-06994-2
Sequence 2, Application PC/TUS9406994
GENERAL INFORMATION:
APPLICANT: The Scripps Research Institute
TITLE OF INVENTION: ANTIBODIES TO HUMAN C5a RECEPTOR
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Spensley Horn Juba & Lubitz
STREET: 1880 Century Park East, Suite 500
CITY: Los Angeles
STATE: California

COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/06994
FILING DATE: 20-JUN-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Tumarkin, Lisa A.
REGISTRATION NUMBER: P-38,347
REFERENCE/DOCKET NUMBER: PD-2790
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Peptide
IMMEDIATE SOURCE:
CLONE: C5AR(133-148)
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..16
PCT-US94-06994-2

Query Match 80.0% Score 28; DB 5; Length 16;
Best Local Similarity 80.0%; Pred. NO. 10;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPMX 5
1111
Db 7 FKPM 11

RESULT 5
US-08-513-974B-351
Sequence 351, Application US/08513974B
Patent No. 6114139
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
APPLICANT: Hosoya, Masaki
APPLICANT: Fujii, Ryo
APPLICANT: Ohtaki, Tetsuya
APPLICANT: Fukusumi, Shoji
APPLICANT: Ohgi, Kazuhiko
TITLE OF INVENTION: G-PROTEIN COUPLED RECEPTOR PROTEIN,
TITLE OF INVENTION: PRODUCTION, AND USE THEREOF
NUMBER OF SEQUENCES: 380
CORRESPONDENCE ADDRESSES:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/513,974B
FILING DATE: 14-SEP-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP95/01599
FILING DATE: 10-AUG-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-093989
FILING DATE: 19-AUG-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-057186
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-007177
FILING DATE: 20-JAN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-326611
FILING DATE: 28-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-270017
FILING DATE: 02-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-236357
FILING DATE: 30-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-236356
FILING DATE: 30-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-189274
FILING DATE: 11-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-189273
FILING DATE: 11-AUG-1945
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-189272
FILING DATE: 11-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Resnick, David S.
REGISTRATION NUMBER: 34,235
REFERENCE/DOCKET NUMBER: 45753
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 351:
SEQUENCE CHARACTERISTICS:
LENGTH: 115 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-513-974B-351

Query Match 80.0% Score 28; DB 3; Length 115;
Best Local Similarity 80.0%; Pred. NO. 72;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPMX 5
1111
Db 4 FKPM 8

RESULT 6
US-08-829-110-3
Sequence 3, Application US/08829110
Patent No. 5882890
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Goli, Surya K.
TITLE OF INVENTION: NOVEL REGULATORS OF G-PROTEIN
TITLE OF INVENTION: SIGNALING
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA

COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/829,110
FILING DATE: Filed Herewith
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0259 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 243 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: THIMNOT02
CLONE: 343504
US-08-829-110-3

Query Match 80.0% Score 28; DB 2; Length 243;
Best Local Similarity 80.0% Pred. No. 1.5e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 KPXWR 6
1111
DB 71 KPAWR 75

RESULT 7
US-08-118-270-35
Sequence 35, Application US/08118270
Patent No. 5508384
GENERAL INFORMATION:
APPLICANT: Murphy, Randall B.
APPLICANT: Schuster, David I.
TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
NUMBER OF SEQUENCES: 348
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/118,270
FILING DATE: 09-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/943,236
FILING DATE: 10-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Townsend, Kevin G.
REGISTRATION NUMBER: 34,033

REFERENCE/DOCKET NUMBER: MURPHY-2A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 304 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-118-270-35

Query Match 80.0% Score 28; DB 1; Length 304;
Best Local Similarity 80.0% Pred. No. 1.9e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPMW 5
1111
DB 103 FKPMW 107

RESULT 8
PCT-US93-08528-35
Sequence 35, Application PC/TUS9308528
GENERAL INFORMATION:
APPLICANT: New York University
TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
NUMBER OF SEQUENCES: 348
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/08528
FILING DATE: 09-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/943,236
FILING DATE: 10-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Townsend, Kevin G.
REGISTRATION NUMBER: 34,033
REFERENCE/DOCKET NUMBER: MURPHY-2 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 304 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US93-08528-35

Query Match 80.0% Score 28; DB 5; Length 304;
Best Local Similarity 80.0% Pred. No. 1.9e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPMW 5

Db 103 FKPIW 107

RESULT 9

US-08-458-970A-9
Sequence 9, Application US/08458970A
Patent No. 5861272
GENERAL INFORMATION:
APPLICANT: LI, ET AL.
TITLE OF INVENTION: C5a Receptor
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,970A
FILING DATE: June 2, 1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/09234
FILING DATE: 16 AUG 1994
ATTORNEY/AGENT INFORMATION:
NAME: MULLINS, J.G.
REGISTRATION NUMBER: 33,073
REFERENCE/DOCKET NUMBER: 325800-353
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 350 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-458-970A-9

Query Match 80.0% Score 28; DB 2; Length 350;
Best Local Similarity 80.0% Pred. No. 2.2e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPIW 5
Db 139 FKPIW 143

RESULT 10

US-09-082-310-1
Sequence 1, Application US/09082310
Patent No. 6096526
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil C.
APPLICANT: Guegler, Karl J.
APPLICANT: Yue, Henry
TITLE OF INVENTION: HUMAN NUCLEIC ACID METHYLASES
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE

CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/082,310
FILING DATE: Herewith
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: CERRONE, MICHAEL C.
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0520 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRSTNOT07
CLONE: 2124957
US-09-082-310-1

Query Match 80.0% Score 28; DB 3; Length 463;
Best Local Similarity 80.0% Pred. No. 2.9e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 KPWR 6
Db 293 KPWR 297

RESULT 11

US-08-876-874-2
Sequence 2, Application US/08876874
Patent No. 5942405
GENERAL INFORMATION:
APPLICANT: Ames, Robert
APPLICANT: Bergsma, Derk
APPLICANT: Foley, James
APPLICANT: Kumar, Chandrika
APPLICANT: Sarau, Henry
TITLE OF INVENTION: THERAPEUTIC AND SCREENING
METHODS USING C3A RECEPTOR AND C3A
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: RATNER & PRESTIA
STREET: P.O. BOX 980
CITY: VALLEY FORGE
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSRO for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,874
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/019,627
FILING DATE: 16-JUN-1997

ATTORNEY/AGENT INFORMATION:

NAME: PRESTIA, PAUL F

REGISTRATION NUMBER: 23,031

REFERENCE/DOCKET NUMBER: P50501

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-407-0700

TELEFAX: 610-407-0701

TELEX: 846169

INFORMATION FOR SEQ. ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 482 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-876-874-2

Query Match

Best Local Similarity 80.0%; Score 28; DB 2; Length 482;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKW 5

Db 125 FKPKW 129

RESULT 12

US-08-906-769-113

Sequence 113, Application US/08906769

Patent No. 6077687

GENERAL INFORMATION:

APPLICANT: Grieve, Robert B.

APPLICANT: Rushlow, Keith E.

APPLICANT: Wu Hunter, Shirley

APPLICANT: Frank, Glenn R.

APPLICANT: Stiegler, Gary

APPLICANT: Gaines, Patrick J.

APPLICANT: Silver, Gary

TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID

TITLE OF INVENTION: MOLECULES AND USES THEREOF

NUMBER OF SEQUENCES: 190

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheridan Ross & McIntosh

STREET: 1700 Lincoln Street, Suite 3500

CITY: Denver

STATE: Colorado

COUNTRY: USA

ZIP: 80203

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/906,769

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/639,075

FILING DATE: 24-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: Connell, Gary J.

REGISTRATION NUMBER: 32,020

REFERENCE/DOCKET NUMBER: 2618-25-C2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 863-9700

TELEFAX: (303) 863-0223

INFORMATION FOR SEQ. ID NO: 113:

SEQUENCE CHARACTERISTICS:

LENGTH: 496 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-906-769-113

Query Match

Best Local Similarity 80.0%; Score 28; DB 3; Length 496;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKW 5

Db 358 FKPKW 362

RESULT 13

US-08-906-616-113

Sequence 113, Application US/08906616

Patent No. 6121035

GENERAL INFORMATION:

APPLICANT: Grieve, Robert B.

APPLICANT: Rushlow, Keith E.

APPLICANT: Wu Hunter, Shirley

APPLICANT: Frank, Glenn R.

APPLICANT: Stiegler, Gary

APPLICANT: Gaines, Patrick J.

APPLICANT: Silver, Gary

TITLE OF INVENTION: FLEA AMINOPEPTIDASE PROTEINS AND USES THEREOF

NUMBER OF SEQUENCES: 190

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheridan Ross P.C.

STREET: 1700 Lincoln Street, Suite 3500

CITY: Denver

STATE: Colorado

COUNTRY: USA

ZIP: 80203

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/906,616

FILING DATE: 05-AUG-1997

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Connell, Gary J.

REGISTRATION NUMBER: 32,020

REFERENCE/DOCKET NUMBER: 2618-25-C2-3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 863-9700

TELEFAX: (303) 863-0223

INFORMATION FOR SEQ. ID NO: 113:

SEQUENCE CHARACTERISTICS:

LENGTH: 496 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-906-616-113

Query Match 80.0%; Score 28; DB 3; Length 496;

Best Local Similarity 80.0%; Pred. No. 3.1e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKW 5

Db 358 FKPKW 362

RESULT 14

US-08-817-795-113

Sequence 113, Application US/08817795

Patent No. 6139840

GENERAL INFORMATION:

APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Hunter, Shirley-Wu
APPLICANT: Frank, Glenn R.
APPLICANT: Heath, Andrew W.
APPLICANT: Yamaka, Miles Yamanaka
APPLICANT: Artsten, Ann
APPLICANT: Dale, Beverly
APPLICANT: Stiegler, Gary
TITLE OF INVENTION: USE OF PROTEASE INHIBITORS AND
TITLE OF INVENTION: PROTEASE VACCINES TO PROTECT ANIMALS FROM FLEA
TITLE OF INVENTION: INFESTATION, AND FLEA PROTEASE PROTEINS, NUCLEIC ACID
TITLE OF INVENTION: MOLECULES, AND USES THEREOF
NUMBER OF SEQUENCES: 119
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/817,795
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/14442
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gary J. Connell
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 113:
SEQUENCE CHARACTERISTICS:
LENGTH: 496 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-817-795-113

Query Match 80.0%; Score 28; DB 4; Length 496;
Best Local Similarity 80.0%; Pred. NO. 3.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 FKPM 5
1111
Db 358 FKPM 362

RESULT 15
US-08-639-075A-113
Sequence 113, Application US/08639075A
Patent No. 6150125
GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Wu Hunter, Shirley
APPLICANT: Frank, Glenn R.
APPLICANT: Stiegler, Gary
APPLICANT: Gaines, Patrick J.
APPLICANT: Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
TITLE OF INVENTION: MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190

CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/639,075A
FILING DATE: 24-Apr-1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 113:
SEQUENCE CHARACTERISTICS:
LENGTH: 496 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-639-075A-113

Query Match 80.0%; Score 28; DB 4; Length 496;
Best Local Similarity 80.0%; Pred. NO. 3.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 FKPM 5
1111
Db 358 FKPM 362

Search completed: February 27, 2002, 11:36:05
Job time: 144 sec

XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 27: SEQ ID No 21652; 487pp; English.

CC The present invention relates to human single exon nucleic acid probes
CC (SENP: see A110068-A128459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 71 AA:

Query Match 90.3%; Score 28; DB 22; Length 71;
Best Local Similarity 66.7%; Pred. No. 77;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
1 1 1 1
1 fhpxmr 6

Db

RESULT 2
AAM29310
ID AAM29310 standard; Protein: 71 AA.

AC AAM29310:
XX
XX 17-OCT-2001 (first entry)
XX
XX Peptide #3347 encoded by probe for measuring placental gene expression.
XX
XX Probe: microarray: human; placenta: antenatal diagnosis:
XX genetic disorder.
XX
XX Homo sapiens.
XX
XX MO200157272-A2.
XX
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR:
XX
XX WPI: 2001-488897/53.
XX
XX human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 27: SEQ ID No 29579; 654pp; English.

CC The present invention relates to single exon nucleic acid probes (SENP:
CC see A1131315-A1157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for

CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.

SO Sequence 71 AA:

Query Match 90.3%; Score 28; DB 22; Length 71;
Best Local Similarity 66.7%; Pred. No. 77;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
1 1 1 1
1 fhpxmr 6

Db

RESULT 3
AAM04538
ID AAM04538 standard; Protein: 71 AA.

AC AAM04538:
XX
XX 09-OCT-2001 (first entry)
XX
XX Peptide #3220 encoded by probe for measuring breast gene expression.
XX
XX Probe: human; breast disease: breast cancer; development disorder:
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX Homo sapiens.
XX
XX MO200157270-A2.
XX
XX 09-AUG-2001.
XX
XX 29-JAN-2001; 2001WO-US00661.
XX
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR:
XX
XX WPI: 2001-476286/51.
XX
XX Novel single exon nucleic acid probe used to measuring gene expression
XX in a human breast -
XX
XX Claim 27: SEQ ID No 13278; 322pp; English.

CC The present invention relates to novel single exon nucleic acid probes
CC (see A1100010-A1110067). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for measuring human gene expression in
CC a human breast sample, where the probe hybridises at high stringency to a
CC nucleic acid expressed in the human breast. The probes are useful for
CC predicting, diagnosing, grading, staging, monitoring and prognosing
CC diseases of the human breast, particularly those diseases with polygenic
CC aetiology. The diseases include: breast cancer, disorders of development,
CC inflammatory diseases of the breast, fibrocystic changes, proliferative
CC breast disease and non-carcinoma tumours.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 71 AA:

Query Match 90.3% Score 28: DB 22: Length 71:
Best Local Similarity 66.7% Pred. No. 77:
Matches 4: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

OY 1 FXPXMR 6
1 1 1 1
1 flnpavr 6

Db

RESULT 4
AAB41139
ID AAB41139 standard: protein: 277 AA.
XX AAB41139:
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF903 polypeptide sequence SEQ ID NO:1806.
XX
KW Human; open reading frame: ORFX; detection: cytosstatic; hepatotropic;
KW vulnerable; antiproliferative; antiparkinsonian; noctropic; neuroprotective;
KW anticonvulsant; osteopathic; antiallergic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotrophic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antineoplastic; antihypertensive;
KW antiviral; antibacterial; antifungal; antihemorrhagic; antihypertensive;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypochyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antineoplastic disease; coagulation;
KW thrombosis; contraceptive.
XX
XX Homo sapiens.
XX
XX PN WO200058473-A2.
XX
XX PD 05-OCT-2000.
XX
XX PF 31-MAR-2000: 2000MO-US08621.
XX
XX PR 31-MAR-1999: 99US-0127607.
XX PR 02-APR-1999: 99US-0127636.
XX PR 05-APR-1999: 99US-0127728.
XX PR 30-MAR-2000: 2000US-0540763.
XX
XX PA (CURA-) CURAGEN CORP.
XX
XX PT Shinkets RA, Leach M:
XX
XX DR WPI: 2000-602362/57.
XX N-PSDB: AAC75348.
XX
XX PT Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 11: Page 1407: 5507pp; English.

AAC74446 to AAC77506 encode the proteins given in AAB40237 to AAB43397,
which represent the human ORFX open reading frames 1 to 3161. The ORFX
sequences have activities such as: cytosstatic; hepatotropic; vulnerable;
antiproliferative; antiparkinsonian; noctropic; neuroprotective;
CC osteopathic; anticonvulsant; antiallergic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotrophic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antineoplastic; antibacterial; antiviral; antifungal; antihemorrhagic;
CC antihypertensive; antianemic. The sequences can be used for determining
the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
vectors. The proteins and nucleic acids may be used to treat cancers.

CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypochyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antineoplastic disease; to enhance
CC coagulation, to inhibit thrombosis; and as a contraceptive.
XX
XX Sequence 277 AA:

Query Match 90.3% Score 28: DB 21: Length 277:
Best Local Similarity 66.7% Pred. No. 2,9e+02:
Matches 4: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

OY 1 FXPXMR 6
1 1 1 1
1 flnpavr 6

Db 271 flnpavr 276

RESULT 5
AAM24361
ID AAM24361 standard: protein: 357 AA.
XX
XX NC AAM24361:
XX
XX DT 12-OCT-2001 (first entry)
XX
DE Human EST encoded protein SEQ ID NO: 1886.
XX
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition.
XX
XX CS Homo sapiens.
XX
XX PN WO200154477-A2.
XX
XX TN 02-AUG-2001.
XX
XX PD 25-JAN-2001: 2001MO-US02687.
XX
XX PF 25-JAN-2000: 2000US-0491404.
XX PR 17-JUL-2000: 2000US-0617746.
XX PR 03-AUG-2000: 2000US-0631451.
XX PR 15-SEP-2000: 2000US-0663870.
XX
XX PA (HYSE-) HYSE-INC.
XX
XX PT Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V:
XX PI Cao Y, Dmanac RA, Zhang J, Werthman T:
XX
XX DR WPI: 2001-476164/51.
XX N-PSDB: AAH99020.
XX
XX PT Isolated polypeptide for treatment of diseases, diagnostics, raising
XX antibodies and research use -
XX
XX Claim 20: Page 1218-1219: 1275pp; English.

The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a
XX protein of the invention.
XX
XX Sequence 357 AA:

Query Match 90.3%; Score 28; DB 22; Length 357;
 Best Local Similarity 66.7%; Pred. No. 3.8e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXMR 6
 1 1 1 1
 Db 241 fsprvr 246

RESULT 6

AAU00428
 ID AAU00428: standard; Protein: 383 AA.

AC AAU00428:

DT 14-MAY-2001 (first entry)

DE Rat Gas1 protein.

XX Rat; Gas1: lethal cell protein; apoptosis; cell death;

KW neurological disorder; cardiovascular disorder; autoimmune disorder;

KW neuroendocrine disorder; oncological disorder; cancer.

XX Ratus norvegicus.

XX Key Location/Qualifiers

FT Misc-difference 167 /note= "Encoded by TCC"

FT Domain 174..279 /note= "Amphipathic alpha-helix region involved in channel activity and in death-inducing property of Gas1"

XX WO200114549-A1.

XX 01-MAR-2001.

PF 21-AUG-2000; 2000MO-EP08182.

XX 24-AUG-1999; 99EP-0306702.

XX (JANC) JANSSEN PHARM NV.

PI Luyten WHML, Naranjo JR, Mellstroem B;

XX WPI: 2001-218449/22.

DR N-PSDB; AAU00428.

PT Inhibiting lethal effect of protein in cell, involves inhibiting
 PT function and/or expression of Gas1 protein and expression of sequence
 PT encoding otherwise lethal protein in the cell

XX Claim 32; Page 58-59; 65pp; English.

XX The present sequence represents rat Gas1 protein, a lethal cell protein.
 CC Novel methods of inhibiting the lethal effect of Gas1 or a functional
 CC equivalent, derivative or bioprecursor which is capable of inducing
 CC apoptosis (AAU00429) involve the use of antisense nucleotide sequences
 CC (AAU00433-AAU00434) which inhibit the expression of the lethal proteins
 CC in a cell. A nucleotide encoding Gas1 or a protein capable of inducing
 CC apoptosis, an antisense molecule, Gas1 polypeptide or fragment of, an
 CC inhibitor or accelerator of cell death, or a pharmaceutical composition
 CC comprising any of these are useful as a medicament for the prevention or
 CC treatment of a disease condition such as neurological disorder
 CC (Parkinson's disease, Alzheimer's disease, Huntington's disease,
 CC amyotrophic lateral sclerosis, or a neurological condition caused by
 CC thrombosis or cerebral trauma), cardiovascular disorder (heart attack),
 CC autoimmune disorder (multiple sclerosis), neuroendocrine disorder
 CC (necrosis of the pituitary gland), or oncological disorder (cancer),
 CC mediated at least in part by expression of Gas1 or its functional
 CC equivalent, derivative or bioprecursor capable of inducing apoptosis in
 CC a cell or a protein in the pathway of which Gas1 is a component. An

CC antisense molecule is useful as a probe, as a medicament or in the
 CC preparation of a drug for treating the above mentioned diseases.
 XX

XX Sequence 383 AA;

Query Match 90.3%; Score 28; DB 22; Length 383;
 Best Local Similarity 66.7%; Pred. No. 4e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXMR 6
 1 1 1 1
 Db 138 fsprvr 143

RESULT 7

AAV36300
 ID AAV36300 standard; Protein: 488 AA.

AC AAV36300:

PT 17-SEP-1999 (first entry)

XX Human secreted protein encoded by gene 77.

XX Human; secreted protein; cancer; tumour; developmental abnormality;

KW foetal deficiency; blood disorder; immune system disorder; inflammation;

KW autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;

KW schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;

KW atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;

XX digestive disorder; endocrine disorder; infection; AIDS.

XX Homo sapiens.

XX WO931117-A1.

XX 24-JUN-1999.

PF 17-DEC-1998; 98WO-US27059.

XX 19-DEC-1997; 97US-0068369.

XX 18-DEC-1997; 97US-0068006.

XX 18-DEC-1997; 97US-0068007.

XX 18-DEC-1997; 97US-0068008.

XX 18-DEC-1997; 97US-0068053.

XX 18-DEC-1997; 97US-0068054.

XX 18-DEC-1997; 97US-0068057.

XX 18-DEC-1997; 97US-0068064.

XX 19-DEC-1997; 97US-0070923.

XX 19-DEC-1997; 97US-0068169.

XX 19-DEC-1997; 97US-0068365.

XX 19-DEC-1997; 97US-0068367.

XX 19-DEC-1997; 97US-0068368.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Carter KC, Duan RD, Feng P, Ferrie AM, Florence C;

XX Florence K, Greene JM, Janat F, Kyaw H, Moore PA;

XX NI J, Rosen CA, Ruben SM, Shi Y, Soppet DR, Wei Y;

XX Yu G;

XX WPI: 1999-418749/35.

XX N-PSDB; AAV36300.

XX New isolated human genes encoding secreted polypeptides

XX Claim 11; Page 387-389; 537pp; English.

XX AAX97916 to AAX98029 represent 110 isolated human secreted protein

XX genes. AAV36224 to AAV36227 represent the secreted proteins encoded by

XX the 110 human genes. The genes and their corresponding secreted

XX polypeptides are useful for preventing, creating or ameliorating medical

XX conditions, e.g. by protein or gene therapy. Also pathological conditions

CC can be diagnosed by determining the amount of the new polypeptides in a
 CC sample or by determining the presence of mutations in the new genes.
 CC Specific uses are described for each of the 110 genes, based on which
 CC tissues they are most highly expressed in, and include developing
 CC products for the diagnosis or treatment of cancer, tumours, developmental
 CC abnormalities and foetal deficiencies, blood disorders, diseases of the
 CC immune system, autoimmune diseases, inflammation, allergies, Alzheimer's
 CC and cognitive disorders, schizophrenia, arthritis, asthma, psoriasis,
 CC sepsis, skin disorders, atherosclerosis, diabetes, cardiovascular
 CC disorders, kidney disorders, digestive/endocrine disorders, infections
 CC and AIDS. The polypeptides are also useful for identifying their binding
 CC partners. The sequences given in AAX97907 to AAX97915 and AAY36223 are
 CC used in the exemplification of the present invention.

XX Sequence 488 AA:

Query Match 90.3%; Score 28; DB 20; Length 488;
 Best Local Similarity 66.7%; Pred. No. 5.1e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 Db 241 fspjwr 246

RESULT 8

AAC92932 standard; Protein: 509 AA.

XX AAC92932:

DT 26-SEP-2001 (first entry)

DE C glutamicum protein fragment SEQ ID NO: 6686.

KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

KM organic acid synthesis.

XX Corynebacterium glutamicum.

PN EPI108790-A2.

PD 20-JUN-2001.

XX 18-DEC-2000; 2000EP-0127688.

XX 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX (KYOWA) KYOWA HAKKO KOGYO KK.

PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

DR Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI: 2001-376931/40.

DR N-PSDB: AAH68151.

PT Novel polynucleotides derived from Coryneform bacteria, for identifying

PT mutation point of a gene, measuring expression of a gene, analysing

PT expression profile or pattern of a gene and identifying homologous gene

PT Claim 17: SEQ ID NO: 6686; 246bp + Sequence Listing: English.

XX The present invention provides a number of nucleotide and protein

CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These

CC are useful for identifying the mutation point of a gene derived from a

CC mutant of Coryneform bacterium, measuring expression amount and

CC analysing the expression profile or expression pattern of a gene derived

CC from Coryneform bacterium, and identifying a homologue of a gene derived

CC from Coryneform bacterium. Coryneform bacteria are useful for producing

CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.

XX Sequence 509 AA:

Query Match 90.3%; Score 28; DB 22; Length 509;
 Best Local Similarity 66.7%; Pred. No. 5.3e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 Db 364 fapjwr 369

RESULT 9

AAB79335 standard; Protein: 509 AA.

XX AAB79335:

DT 30-APR-2001 (first entry)

DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:186.

KW Corynebacterium glutamicum; carbon metabolism and energy production;

KW SMP protein; sugar metabolism and oxidative phosphorylation protein;

KW fine chemical production; organic acid; proteinogenic amino acid;

KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;

KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;

KW carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;

KW diagnosis; Corynebacterium diptheriae; evolutionary study.

XX Corynebacterium glutamicum.

OS WO200100844-A2.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000WO-1800943.

XX 25-JUN-1999; 99US-0141031.

XX 08-JUL-1999; 99DE-1031412.

XX 08-JUL-1999; 99DE-1031413.

XX 08-JUL-1999; 99DE-1031419.

XX 08-JUL-1999; 99DE-1031420.

XX 08-JUL-1999; 99DE-1031424.

XX 08-JUL-1999; 99DE-1031428.

XX 08-JUL-1999; 99DE-1031431.

XX 08-JUL-1999; 99DE-1031433.

XX 08-JUL-1999; 99DE-1031434.

XX 08-JUL-1999; 99DE-1031510.

XX 08-JUL-1999; 99DE-1031562.

XX 08-JUL-1999; 99DE-1031634.

XX 09-JUL-1999; 99DE-1032180.

XX 09-JUL-1999; 99DE-1032227.

XX 09-JUL-1999; 99DE-1032230.

XX 09-JUL-1999; 99US-0143208.

XX 14-JUL-1999; 99DE-1032924.

XX 14-JUL-1999; 99DE-1032973.

XX 14-JUL-1999; 99DE-1033005.

XX 27-AUG-1999; 99DE-1040765.

XX 31-AUG-1999; 99US-0151572.

XX 03-SEP-1999; 99DE-1042076.

XX 03-SEP-1999; 99DE-1042079.

XX 03-SEP-1999; 99DE-1042086.

XX 03-SEP-1999; 99DE-1042087.

XX 03-SEP-1999; 99DE-1042088.

XX 03-SEP-1999; 99DE-1042095.

PR	03-SEP-1999:	99DE-1042123.
PR	03-SEP-1999:	99DE-1042125.
XX	(BADI) BASF AG.	
PA	Pompejus M., Kroegeer B, Schroeder H, Zelder O, Haberbauer G;	
PI	N-PsDB; AAF71452.	
XX	WPI: 2001-061975/07.	
DR	New isolated Corynebacterium glutamicum nucleic acid encoding a sugar	
XX	metabolism and oxidative phosphorylation protein for production or	
PT	modulation of production of fine chemicals e.g. amino acids,	
PT	carbohydrates or enzymes -	
XX	Clalm 20; Page 407-409; 1246pp; English.	
PS	AAF71360 to AAF71750 encode the corynebacterium glutamicum sugar	
XX	metabolism and oxidative phosphorylation (SMP) proteins given in	
CC	AAB79243 to AAB 79633 which are involved in carbon metabolism and	
CC	energy production. The C. glutamicum SMP gene can be used in vectors	
CC	(ii) for expression in host cells and production or modulation of	
CC	production of fine chemicals, such as, an organic acid, a proteinoenic	
CC	or nonproteinoenic amino acid (preferred), a purine or pyrimidine base,	
CC	a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty	
CC	acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a	
CC	cofactor, a polypeptide, or an enzyme. The presence of (I) or SMP proteins	
CC	(II) encoded by them are used for diagnosing the presence or activity of	
CC	Corynebacterium diptheriae in a subject. (I), (II), (III) or host cells	
CC	containing them are used to map genomes of organisms related to	
CC	C. glutamicum, identify and localise C. glutamicum sequences of interest,	
CC	In evolutionary studies, in determining SMP protein regions required	
CC	for function, in modulating SMP protein activity, in modulating the	
CC	metabolism of sugars, and in modulating high-energy molecule production .	
CC	In a cell (I.e. ATP, NADPH).	
XX		
SQ	Sequence	509 AA:
OY	1 FXPXMR 6	
Db	364 fapxvr 369	
RESULT 10		
ID	AAW75776	
XX	AAW75776 standard; Protein: 527 AA.	
XX	AAW75776:	
XX	21-DEC-1998 (first entry)	
DE	Mycobacterium tuberculosis cellular uptake Mcsp protein.	
KM	Mycobacterium cell entry; Mcsp: cellular uptake; tuberculosis;	
XX	Infection; vaccine; therapy; diagnosis; antibody; macrophage.	
OS	Mycobacterium tuberculosis strain H37Ra (ATCC 25177).	
FH	Key	Location/Qualifiers
FT	Misc-difference 66	
FT	Misc-difference 202	/note= "encoded by AT"
FT	Misc-difference 202	/note= "encoded by CG"
XX	WO9840098-A1.	
XX	17-SEP-1998.	
XX		

Pf		09-MAR-1998;	98WO-USO4653.
Xx			
Pt		06-AUG-1997;	97US-0907229.
Pt		10-MAR-1997;	97US-0040097.
XX		(CORR) CORNELL RES FOUND INC.	
PA			
xx			
Riley LM:			
DR		WPI: 1998-506483/43.	
N-PSDB; AAV52613.			
New isolated protein from Mycobacterium tuberculosis - useful for e.g. diagnosis, treatment or prevention of infection and for mediating uptake of materials by mammalian cells			
Claim 9; Page 12-13; 62pp; English.			
This is the deduced amino acid sequence of a protein encoded by a DNA molecule (see AAU52613) associated with conferring on Mycobacterium tuberculosis an ability to enter mammalian cells. The protein, termed Mcpep, or its fragments, especially polypeptides comprising amino acids 1-167, 95-167, 95-528 or 319-528 of the protein, can be expressed in host cells for recombinant production. The proteins are useful in vaccines to prevent infection by M. tuberculosis, and antibodies raised against the proteins can be used for passive immunisation. The proteins, antibodies and DNA molecules may be utilised in diagnostic assays to detect M. tuberculosis in tissue or bodily fluids. The protein can be associated with various other therapeutic materials (e.g. antibiotics, DNA fragments or anti-neoplastic agents) for administration to mammals, particularly humans, to achieve cellular uptake of those materials. The ability of Mcpep to augment uptake into macrophages provides an opportunity to deliver genes specifically to macrophages to induce humoral and cell-mediated immunity.			
Sequence 527 AA;			

```

Query March 90.3% Score 28 DB 19 Length 527;
Best Local Similarity 66.7% Pred. NO. 5.5e+02;
Matches 4: Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 EXPXMR 6
   | | |
   | | |
Db 3 fgspr 8

=====
::RESULT 11
ID AAM48359 standard; Protein: 527 AA.
XX
XX AAM48359;
AC
XX
XX
XX 03-JUL-1994 (first entry)
XX
XX Mycobacterium bovis BCG protein.
XX
XX Mycobacterium bovis BCG protein; Mcep; Tuberculosis; BCGINV-33;
XX hybridisation test; amplification; probe.
XX
XX Mycobacterium bovis.
XX
XX W09801559-A1.
XX
XX 15-JAN-1998.
XX
XX
XX 09-JUL-1997; 97WO-CA00484.
XX
XX 10-JUL-1996; 96US-0677970;
XX
XX (CONN-) CONNAUGHT LAB LTD.

```

```

XX  Anand NN, Klein MH;
PI  WPI: 1998-110232/10.
DR  N-PSDB: AAV17758.
XX
PT  Nucleic acid encoding mycobacterial protein involved in cell binding
PT  and entry - used for diagnosis of Mycobacterium infection and in
PT  vaccines for humans or animals
XX
PS  Claim 8: Fig 7A-7E: 107bp; English.
XX
CC  The present-sequence represents a Mycobacterial bovis BCG protein
CC  (45-60 kDa) associated with cell binding and entry. The BCG protein is
CC  encoded by the BCGINV-33 gene whose sequence contains two translational
CC  frames, one for the M.bovis BCGINV-33 gene and the other for the
CC  M.tuberculosis Mce gene. Therefore the Mcep protein is an internal
CC  polypeptide of the BCG protein from M.bovis. The invention relates to
CC  the use of the BCGINV-33 gene sequence in hybridisation tests for
CC  diagnosis of Mycobacterial infections, especially M.tuberculosis
CC  infections. Fragments of the BCGINV-33 gene sequence (AAV17758) can be
CC  used in PCR to detect Mycobacterium in tissues and body fluids and also
CC  for isolating related genes. BCG protein or its fragments can be used
CC  in vaccines to generate an immune response that may be capable of
CC  protecting humans and animals (especially cattle) against mycobacterial
CC  infections.
XX
SO  Sequence 527 AA:

Query Match          90.3%; Score 28; DB 19; Length 527;
Best Local Similarity 66.7%; Pred. No. 5.5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  1 FXPXWR 6
    | | | |
Db  3 fypswr 8

RESULT 12
AAV35425
ID  AAV35425 standard; Protein: 629 AA.
XX
AC  AAV35425;
XX
DT  13-SEP-1999 (first entry)
XX
DE  Protein involved in transcription, translation and/or maturation.
XX
KW  Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW  sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW  vaccine; neutralising epitope.
XX
OS  Chlamydia pneumoniae.
XX
PN  MO9927105-A2.
XX
PD  03-JUN-1999.
XX
PF  20-NOV-1998; 98WO-1801890.
XX
PR  04-NOV-1998; 98US-0107078.
PR  21-NOV-1997; 97FR-0014673.
XX
PA  (CEST ) GENSET.
XX
PI  Griffiths R;
XX
DR  WPI: 1999-357842/30.
XX
PT  Genome sequence of Chlamydia pneumoniae
XX
PS  Page 1207-1209; Disclosure: 1912pp; English.

```

```

XX  AAV34584-Y35879 represent the proteins encoded by all the open reading
CC  frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
CC  C. pneumoniae causes respiratory disease such as pneumonia and
CC  bronchitis and is thought to be a contributing factor in heart
CC  disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC  nodosum or pharyngitis. The polypeptides encoded by the open reading
CC  frames of the C. pneumoniae genome (see AAV34584-Y35879) can be used in
CC  immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC  nucleotide sequences can also be used as immunogenic compositions,
CC  especially where the vector directs the expression of a neutralising
CC  epitope of C. pneumoniae.
XX
SO  Sequence 629 AA:

Query Match          90.3%; Score 28; DB 20; Length 629;
Best Local Similarity 66.7%; Pred. No. 6.6e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY  1 FXPXWR 6
    | | | |
Db  290 faplw 295

RESULT 13
AAE03765
ID  AAE03765 standard; Protein: 766 AA.
XX
AC  AAE03765;
XX
DT  07-AUG-2001 (first entry)
XX
DE  Human gene 2 encoded secreted protein HCE3C63, SEQ ID NO:35.
XX
KW  human; secreted protein; proliferative disorder; cancer; tumour;
KW  foetal abnormality; developmental abnormality; haematopoietic disorder;
KW  immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW  inflammation; allergy; neurological disorder; Alzheimer's disease;
KW  Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW  skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW  cardiovascular disorder; angiogenic disorder; kidney disorder;
KW  gastrointestinal disorder; pregnancy-related disorder; gene therapy;
KW  endocrine disorder; infection; wound healing; vulvar;
KW  cell culture; chemotaxis; food additive;
KW  binding partner identification.
XX
OS  Homo sapiens.
XX
FT  Key Location/Qualifiers
FT  Peptide 1..33
FT  Protein /label= signal_peptide
FT  34..766
FT  /note= "Mature secreted protein"
XX
PN  MO200132837-A1.
XX
PD  10-MAY-2001.
XX
PF  17-OCT-2000; 2000MO-US28664.
XX
PR  02-NOV-1999; 99US-0163085.
PR  17-DEC-1999; 99US-0172411.
XX
PA  (HUMA-) HUMAN GENOME SCI INC.
XX
PI  Fischella M, Wei P, Lafleur DW, Olsen HS, Baker K, Ebner R;
PI  Komatsoulis G, Rosen CA, Ruben SM, Duan RD, Young PE, Florence KA;
PI  Moore PA, Birse CE, Ni J, Soppet DR, Shi Y;
XX
DR  WPI: 2001-328782/34.
XX
PT  N-PSDB: AAD08192.
XX

```

PT Novel human secreted proteins and nucleic acids for diagnosing,
PT preventing and treating neurological, cardiovascular, infectious,
PT autoimmune, gastrointestinal, bone disorders, cancer, particularly
PT ovarian cancer -
XX
XX
PS Claim 11: Page 387-390; 421pp; English.
XX
XX AAD08191-AAD08213 represent cDNAs corresponding to 19 human secreted
CC protein genes and AAE03764-AAE03786 represent the proteins they encode.
CC AAE03787-AAE03800 represent human secreted protein fragments or variants.
CC The genes and their secreted proteins are useful for preventing,
CC treating, or ameliorating medical conditions, e.g., by protein or gene
CC therapy; pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 19 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein of the invention.
XX
XX Sequence 766 AA:
QY
Db 525 fdpswr 530
1 fpxpwr 6
1 1 1 1
1 1 1 1
1 1 1 1

Query Match 90.3%; Score 28; DB 22; Length 766;
Best Local Similarity 66.7%; Pred. No. 8e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Caps 0;

RESULT 14
ID AAB95425 standard; Protein: 949 AA.
XX
AC AAB95425:
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:17833.
XX
KW Human: primer: detection; diagnosis; antiscense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EPI074617-A2.
XX
PD 07-FEB-2001.
XX
PE 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPL: 2001-318749/34.
XX
XX
PS Claim 8: SEQ ID 17833; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX Sequence 949 AA:
QY
Db 241 fspjwr 246
1 fpxpwr 6
1 1 1 1
1 1 1 1
1 1 1 1

Query Match 90.3%; Score 28; DB 22; Length 949;
Best Local Similarity 66.7%; Pred. No. 9.8e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Caps 0;

RESULT 15
ID AAM27126 standard; Protein: 2247 AA.
XX
AC AAM27126:
XX
DT 11-FEB-1998 (first entry)
XX
DE Equine rhinovirus 1 (ERHV1) polypeptide amino acid sequence.
XX
KW Equine rhinovirus 1; ERHV1; foot-and-mouth disease virus; vaccine; horse;
KW diagnosis; antigens; polypeptide; enzyme-linked immunosorbent assay;
KW recombinant protein.
XX
OS Equine rhinovirus 1.
XX
US
XX
FH Key Location/Qualifiers
FT Cleavage-site 207..208
FT Cleavage-site 287..288
FT Cleavage-site 517..518
FT Cleavage-site 743..744
FT Cleavage-site 991..992

Query Match 93.5%: Score 29; DB 2; Length 477;
 Best Local Similarity 66.7%: Pred. No. 17e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPXMR 6
 1 1 1 1
 Db 388 FSPSWR 393

RESULT 3
 578224
 virulence-associated protein mKfB - Salmonella typhimurium plasmid
 C:Species: Salmonella typhimurium
 C:Date: 29-Jan-1998 #sequence_revision 13-Feb-1998 #text_change 20-Aug-1999
 C:Accession: S78224; S23713; A43996
 R:Norrel, F.
 Submitted to the EMBL Data Library, January 1992
 A:Reference number: S78224
 A:Accession: S78224
 A:Molecule type: DNA
 A:Residues: 1-563 <NOR>
 A:Cross-references: EMBL:X57096; NID:946998; PIDN:CAA40380.1; PID:946999
 R:Norrel, F.; Plisano, M.R.; Nicoli, J.; Popoff, M.Y.
 Res. Microbiol. 140, 455-457, 1989
 A:Title: Nucleotide sequence of the plasmid-borne virulence gene mKfB from Salmonella ty
 A:Reference number: A43996; MUID:90161559
 A:Accession: S23713
 A:Molecule type: DNA
 A:Residues: 1-511; NILEKHKGRILCDVAHFKGEVEMLF, 514-563 <NOW>
 A:Cross-references: EMBL:X57096
 C:Genetics:
 A:Genome: Plasmid
 A:Superfamily: virulence-associated protein spvB
 C:Keywords: DNA binding; transcription regulation

Query Match 93.5%: Score 29; DB 2; Length 563;
 Best Local Similarity 66.7%: Pred. No. 2e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPXMR 6
 1 1 1 1
 Db 50 FAPAMR 55

RESULT 4
 MW070
 70K protein - potato leaf roll virus (strain 1)
 C:Species: potato leaf roll virus
 A:Note: host Solanum tuberosum (potato)
 C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 21-Jul-2000
 C:Accession: JA0118; S24591
 R:Mayo, M.A.; Robinson, D.J.; Jolly, C.A.; Hyman, L.
 J. Gen. Virol. 70, 1037-1051, 1989
 A:Title: Nucleotide sequence of potato leafroll luteovirus RNA.
 A:Reference number: JA0119; MUID:89279282
 A:Accession: JA0118
 A:Molecule type: genomic RNA
 A:Residues: 1-639 <MAY>
 A:Cross-references: EMBL:X14600; NID:9222293; PIDN:BA00417.1; PID:9222297
 A:Note: the nucleotide sequence was submitted to the EMBL data library, March 1989
 C:Comment: The genome is a single-stranded, positive-sense RNA.
 C:Superfamily: potato leaf roll virus 70K protein

Query Match 93.5%: Score 29; DB 1; Length 639;
 Best Local Similarity 66.7%: Pred. No. 2.2e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPXMR 6
 1 1 1 1

Db 615 FSPSWR 620

RESULT 5
 F82578
 hypothetical protein XF2271 [imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: E82578
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MUID:20365717
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: E82578
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-135 <STM>
 A:Cross-references: GB:AE004039; GB:AE003849; NID:9107425; PIDN:AAF85070.1; GSPDB:GN
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
 briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincant, A.P.; Ferreira, A.J.S.
 Submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraaga, J.S.; Franca, S.C.; Franco, M.C.; Fr
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
 Chido, M.A.; Madalira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.; M.; Mitacca, R.C.; Miyaki, C.
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.G.; Palmieri, L.
 Iodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
 M.; Tshukako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
 A:Reference number: A59328
 A:Contents: annotation
 A:Genetics:
 A:Gene: XF2271

Query Match 90.3%: Score 28; DB 2; Length 135;
 Best Local Similarity 66.7%: Pred. No. 82;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPXMR 6
 1 1 1 1
 Db 79 FVPAMR 84

RESULT 6
 VM03BV
 envelope protein E - Berne virus
 C:Species: Berne virus
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 16-Jul-1999
 C:Accession: A39989; S15570
 R:den Boon, J.A.; Slijder, E.J.; Locker, J.K.; Horzinek, M.C.; Rotzler, P.J.M.
 Virology 182, 655-663, 1991
 A:Title: Another triple-spanning envelope protein among intracellularly budding RNA v
 A:Reference number: A39989; MUID:91220715
 A:Accession: A39989
 A:Molecule type: genomic RNA
 A:Residues: 1-233 <DEN>
 A:Cross-references: GB:X52505; NID:958768; PIDN:CAA36747.1; PID:958769
 R:Boon den, J.A.; Slijder, E.J.; Horzinek, M.C.; Rotzler, P.J.M.
 submitted to the EMBL data library, February 1990
 A:Reference number: S15570
 A:Accession: S15570
 A:Molecule type: preliminary
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-233 <BOO>
 A:Cross-references: EMBL:X52505; NID:958768; PIDN:CAA36747.1; PID:958769
 C:Superfamily: Berne virus envelope protein E
 C:Keywords: envelope protein

Query Match 90.3% Score 28; DB 1; Length 233;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
1 1 1 1
Db 60 FSPXMR 65

RESULT 7

S10929
Transposase - Mycobacterium fortuitum insertion sequence IS6100
C:Species: Mycobacterium fortuitum
A:Variety: strain FCI
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 24-Sep-1999
C:Accession: S10929
R:Martin, C.; Timm, J.; Razzier, J.; Gomez-Lus, R.; Davies, J.; Gicquel, B.
Nature 345: 739-743, 1990
A:Title: Transposition of an antibiotic resistance element in mycobacteria.
A:Reference number: S10927; MUID:90294910
A:Accession: S10929
A:Molecule type: DNA
A:Residues: 1-254 <NAT>
A:Cross-references: EMBL:X53635; NID:944283; PIDN:CAAJ37685.1; PID:944285
C:Genetics:
A:Experimental source: strain FCI
C:Genetics:
A:Mobile element: insertion sequence IS6100
C:Superfamily: Salmonella typhimurium conserved hypothetical protein
C:Keywords: DNA binding

Query Match 90.3% Score 28; DB 2; Length 254;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
1 1 1 1
Db 70 FSPXMR 75

RESULT 8

A49895
Transposase - Flavobacterium sp. insertion sequence IS6100
C:Species: Flavobacterium sp.
C:Date: 11-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 04-Sep-1998
C:Accession: A49895
R:Kato, K.; Ohnuki, K.; Mitsuda, H.; Yomo, T.; Negoro, S.; Urabe, T.
J. Bacteriol. 176: 1197-1200, 1994
A:Title: Insertion sequence IS6100 on plasmid pOAD2, which degrades nylon oligomers.
A:Reference number: A49895; MUID:94148782
A:Accession: A49895
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-254 <KAT>
A:Cross-references: GB:D26094
A:Experimental source: strain K1723T1; plasmid pOAD2; insertion sequence IS6100
C:Genetics:
A:Note: nucleotide sequence not given
A:Mobile element: insertion sequence IS6100
C:Superfamily: Salmonella typhimurium conserved hypothetical protein
C:Keywords: DNA binding

Query Match 90.3% Score 28; DB 2; Length 254;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
1 1 1 1
Db 70 FSPXMR 75

RESULT 9

E84788
Hypothetical protein At2g37100 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84788
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402: 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: E84788
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-297 <STO>
A:Cross-references: GB:AE002093; NID:94371291; PIDN:AAD18149.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g37100
A:Map position: 2

Query Match 90.3% Score 28; DB 2; Length 297;
Best Local Similarity 66.7%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
1 1 1 1
Db 118 FSPXMR 123

RESULT 10

T32776
Hypothetical protein D1069.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 04-Mar-2000
C:Accession: T32776
R:Murray, J.; Langston, Y.; Clarke, K.; Morris, M.
submitted to the EMBL data library, December 1997
A:Description: The sequence of C. elegans cosmid D1069.
A:Reference number: Z21223
A:Accession: T32776
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-311 <MUR>
A:Cross-references: EMBL:AF040641; PIDN:AA894946.1; GSPDB:GN00020; CESP:D1069.1
A:Experimental source: strain Bristol N2; clone D1069
C:Genetics:
A:Gene: CESP:D1069.1
A:Map position: 2
A:Introns: 264/3
C:Superfamily: Caenorhabditis elegans hypothetical protein D1069.1

Query Match 90.3% Score 28; DB 2; Length 311;
Best Local Similarity 66.7%; Pred. No. 1.8e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
1 1 1 1
Db 25 FSPXMR 30

RESULT 11

B84335
Hypothetical protein Vng1839h [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: B84335
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky
; Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
Jung, K.H.; Alam, M.; Freltas, T.
Proc. Natl. Acad. Sci. U.S.A. 97: 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.;

A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483
A:Accession: B84335
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337 <STO>
A:Cross-references: GB:AE004437; NID:g10581284; PIDN:AAG20046.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNC1833H

Query Match 90.3%; Score 28; DB 2; Length 337;
Best Local Similarity 66.7%; Pred. No. 1.9e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
1 1 1 1
Db 79 FDPXMR 84

RESULT 12
C83317
conserved hypothetical protein PA2630 [Imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: C83317
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapid, K.; Lm,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337
A:Accession: C83317
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-389 <STO>
A:Cross-references: GB:AE004691; GB:AE004091; NID:g9948688; PIDN:AAG06018.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA2630

Query Match 90.3%; Score 28; DB 2; Length 389;
Best Local Similarity 66.7%; Pred. No. 2.2e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
1 1 1 1
Db 111 FLPSWR 116

RESULT 13
S18157
globulin 2 - eastern white pine
C:Species: Pinus strobus (eastern white pine)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Aug-1999
C:Accession: S18157
R:Rugh, C.L.; Kamalay, J.C.
submitted to the EMBL Data Library, November 1991
A:Description: Legumin mRNAs from Pinus strobus L.
A:Reference number: S18156
A:Accession: S18157
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-410 <RUC>
A:Cross-references: EMBL:Z11487; NID:g20744; PIDN:CAA77569.1; PID:g20745
C:Superfamily: glycinn

Query Match 90.3%; Score 28; DB 2; Length 410;
Best Local Similarity 66.7%; Pred. No. 2.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
1 1 1 1
Db 269 FVPSWR 274

RESULT 14
A69149
O-antigen transporter related protein - Methanobacterium thermoautotrophicum (strain
C:Species: Methanobacterium thermoautotrophicum
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: A69149
R:Smith, D.R.; Doucette-Stamm, L.A.; Delouche, C.; Lee, H.; Dubois, J.; Aldredge, T
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, K.; Jiwani,
li, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
A:Reference number: A69000; MUID:98037514
A:Accession: A69149
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-475 <MTH>
A:Cross-references: GB:AE000823; GB:AE000666; NID:g2621432; PIDN:AAB84885.1; PID:g2742
C:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH379
A:Start codon: GTC
C:Superfamily: succinoglycan biosynthesis transport protein

Query Match 90.3%; Score 28; DB 1; Length 475;
Best Local Similarity 66.7%; Pred. No. 2.6e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
1 1 1 1
Db 408 FGPAWR 413

RESULT 15
S18156
globulin 1 - eastern white pine
C:Species: Pinus strobus (eastern white pine)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Aug-1999
C:Accession: S18156
R:Rugh, C.L.; Kamalay, J.C.
submitted to the EMBL Data Library, November 1991
A:Description: Legumin mRNAs from Pinus strobus L.
A:Reference number: S18156
A:Accession: S18156
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-409 <RUC>
A:Cross-references: EMBL:Z11486; NID:g20742; PIDN:CAA77568.1; PID:g20743
C:Superfamily: glycinn

Query Match 90.3%; Score 28; DB 2; Length 488;
Best Local Similarity 66.7%; Pred. No. 2.7e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
1 1 1 1
Db 347 FVPSWR 352

Search completed: February 27, 2002, 11:45:21
Job time: 700 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:42:46 : Search time 78.39 Seconds
(without alignments)
2.806 Million cell updates/sec

Title: US-09-446-109A-12
Perfect score: 31
Sequence: 1 FXPXMR 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	93.5	452	1	SPS2_MOUSE
2	29	93.5	477	1	Y264_SYNY3
3	29	93.5	639	1	V70K_PLYV1
4	28	90.3	151	1	HSPD_BRAJA
5	28	90.3	151	1	HSPH_BRAJA
6	28	90.3	153	1	HSPB_BRAJA
7	28	90.3	233	1	VENV_BEV
8	28	90.3	333	1	DPSD_CAEEL
9	28	90.3	508	1	GLPK_MYCLE
10	28	90.3	517	1	GLPK_MYCTU
11	28	90.3	639	1	V70K_PLYVW
12	27	87.1	199	1	YAF7_CAEEL
13	27	87.1	277	1	HO_FUGRU
14	27	87.1	456	1	BES1_YEAST
15	27	87.1	459	1	RBL2_RHOSH
16	27	87.1	514	1	GSN1_BRAJU
17	27	87.1	522	1	GSN1_ARATH
18	27	87.1	535	1	YHIL_ECOLI
19	27	87.1	591	1	IF37_ARATH
20	27	87.1	623	1	PRR2_CANAL
21	27	87.1	644	1	VPA_BTV2A
22	27	87.1	878	1	C105_MOUSE
23	27	87.1	897	1	C105_HUMAN
24	27	87.1	1023	1	TSCC_PSEAM
25	27	87.1	1204	1	RRPL_NDVB
26	26	83.9	224	1	RPIA_LACIA
27	26	83.9	407	1	YK67_MYCTU
28	26	83.9	486	1	LIP1_YARLI
29	26	83.9	575	1	TUCA_ECOLI
30	26	83.9	700	1	YHFK_ECOLI
31	26	83.9	746	1	TAGE_BACSU
32	26	83.9	855	1	YB29_YEAST
33	26	83.9	3206	1	POLG_PSBMV

34	25	80.6	284	1	YAT1_SYNP1	Q05380 synechococ
35	25	80.6	307	1	CRFB_SYNP7	P37269 synechococ
36	25	80.6	493	1	YEB6_YEAST	P39997 saccharomyc
37	25	80.6	497	1	CPD6_HUMAN	P10635 homo sapien
38	25	80.6	497	1	CPDH_MACFA	Q29488 macaca fasc
39	25	80.6	497	1	CPDJ_CALJA	O18992 callithrix
40	25	80.6	500	1	CPDJ_RAT	P13108 rattus norv
41	25	80.6	500	1	CPDG_CAVPO	O64403 cavia porce
42	25	80.6	500	1	CPDI_RAT	O64680 rattus norv
43	25	80.6	523	1	NQ4M_PROWT	Q37617 proteoltheca
44	25	80.6	628	1	YAXN_RHISN	P55706 rhizobium s
45	25	80.6	743	1	ANAG_HUMAN	P54802 homo sapien

ALIGNMENTS

RESULT	ID	SPS2_MOUSE	STANDARD:	PRT:	452 AA.
AC	P97364:				
DT	15-JUL-1999 (Rel. 38, Created)				
DT	15-JUL-1999 (Rel. 38, Last sequence update)				
DT	30-MAY-2000 (Rel. 39, Last annotation update)				
DE	SELENIDE, WATER DIKINASE 2 (EC 2.7.9.3) (SELENOPHOSPHATE SYNTHETASE 2)				
DE	(SELENIUM DONOR PROTEIN 2).				
GN	SPS2.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
CX	NCBI_TaxID=10090;				
PN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=96017645; PubMed=758067;				
RA	Guimaraes M.J., Bazan J.F., Zlotnick A., Wiles M.V., Grimaldi J.C.,				
RA	Lee F., McClanahan T.;				
RT	"A new approach to the study of haematopoietic development in the				
RT	yolk sac and embryoid bodies. "				
RL	Development 121:3335-3346(1995).				
PN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=97140286; PubMed=8986768;				
RA	Guimaraes M.J., Peterson D., Vicari A., Cocks B.C., Copeland N.C.,				
RA	Gilbert D.J., Jenkins N.A., Ferrick D.A., Kastlein R., Bazan J.F.,				
RA	Zlotnick A.;				
RT	"Identification of a novel seld homolog from eukaryotes, bacteria,				
RT	and archaea: is there an autoregulatory mechanism in selenocysteine				
RT	metabolism?";				
FL	Proc. Natl. Acad. Sci. U.S.A. 93:15086-15091(1996).				
CC	- FUNCTION: SYNTHESIZES SELENOPHOSPHATE FROM SELENIDE AND ATP.				
CC	- CATALYTIC ACTIVITY: ATP + SELENIDE + H(2)O = AMP + SELENOPHOSPHATE				
CC	+ PHOSPHATE.				
CC	- COFACTOR: SELENOCYSTEINE. THE ACTIVE-SITE SELENOCYSTEINE IS				
CC	ENCODED BY THE OPAL CODON, UGA.				
CC	- SIMILARITY: BELONGS TO THE SELENOPHOSPHATE SYNTHETASE 1 FAMILY.				
CC	CLASS I SUBFAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL: U43285; AAC53024.1;				
DR	MOP: MGI:108388; SPS2				
DR	InterPro: IPR000726; ATRS-related.				
DR	Pfam: PF00586; ATRS: 1.				
KW	Transferase; Selenium; Selenocysteine; ATP-binding.				
FT	ACT SITE 63 63				
FT	SE CYS 63 63				
FT	SITE 66 66				
FT	IMPORTANT FOR CATALYTIC ACTIVITY (BY				

FT NE-BIND 322 328 SIMILARITY.
 FT DOMAIN 2 9 ATP (POTENTIAL).
 FT DOMAIN 433 440 POLY-ALA.
 SO SEQUENCE 452 AA: 47786 MW: 9DA6F7250CFE80E4 CRC64:

Query Match 93.5%: Score 29; DB 1; Length 452;
 Best Local Similarity 66.7%: Pred. No. 79;
 Matches 4: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 1 1 1 1
 Db 48 FSPSMR 53

RESULT 2
 Y264_SYNY3 STANDARD: PRT: 477 AA.
 AC P73436;
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL 54.0 KDA PROTEIN SLL1464.
 GN SLL1464.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN 111
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hirosewa M., Sugita M., Sasamoto S., Kimura T.,
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Natuo K.,
 RA Okumura S., Shimpō S., Takeuchi C., Wada T., Watanabe A.,
 RA Yamada M., Yasuda M., Tabata S.;
 RT *Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.*;
 RL DNA Res. 3:109-136(1996).
 CC -1- SIMILARITY: BELONGS TO THE UPF0061 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D90906: BAA17476.1;
 DR InterPro: IPR003846: UPF0061.
 DR Pfam: PF02696: UPF0061.1
 KW Hypothetical protein: Complete proteome.
 SO SEQUENCE 477 AA: 54041 MW: 81F6899B1A6D13C CRC64:

Query Match 93.5%: Score 29; DB 1; Length 477;
 Best Local Similarity 66.7%: Pred. No. 83;
 Matches 4: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 1 1 1 1
 Db 388 FSPSMR 393

RESULT 3
 V70K_PLRV1 STANDARD: PRT: 639 AA.
 ID V70K_PLRV1
 AC P17519;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE 69.7 KDA PROTEIN (ORF 2).
 OS Potato leafroll virus (strain 1) (PLRV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
 OC Poterovirus.
 OX NCBI_TaxID=12046;
 RN 111
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89279282; PubMed=2732710;
 RA Mayo M.A., Robinson D.J., Jolly C.A., Hyman L.;
 RT *Nucleotide sequence of potato leafroll luteovirus RNA.*;
 RL J. Gen. Virol. 70:1037-1051(1989).
 CC -1- SIMILARITY: 54% SIMILARITY TO BEET WESTERN YELLOW VIRUS ORF2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D00530: BAA00417.1;
 DR EMBL: X14600: CAA32741.1;
 DR PIR: JAO118; WW070.
 DR PIR: S24591; S24591.
 DR InterPro: IPR000382; Luteo_ORF2.
 DR Pfam: PF02122; Luteo_ORF2.1.
 DR PRINTS: PR00913; LVIRSORF2.
 SO SEQUENCE 639 AA: 69629 MW: AFCE2FB393BEE097 CRC64:

Query Match 93.5%: Score 29; DB 1; Length 639;
 Best Local Similarity 66.7%: Pred. No. 11e+02;
 Matches 4: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 1 1 1 1
 Db 615 FSPSMR 620

RESULT 4
 HSPD_BRAJA STANDARD: PRT: 151 AA.
 ID HSPD_BRAJA
 AC 069241;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE SMALL HEAT SHOCK PROTEIN HSPD.
 GN HSPD.
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Bradyrhizobium group; Bradyrhizobium.
 OX NCBI_TaxID=375;
 RN 111
 RP SEQUENCE FROM N.A.
 RA Narberhaus F., Weighofer W., Fischer H.M., Hennecke H.;
 RT *Identification of the Bradyrhizobium japonicum desp gene as part of
 RT an operon containing small heat shock protein genes.*;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AJ003064: CAA05835.1;
 DR InterPro: IPR002068; Crystallin_HSP20.
 DR Pfam: PF00011; HSP20; 1.

DR PROSITE: PS01031: HSP20: 1.
 KM Heat shock: Multigene family.
 SQ SEQUENCE 151 AA: 17272 MW: FB44EF94FB599EE4 CRC64:

Query Match 90.3%; Score 28; DB 1; Length 151;
 Best Local Similarity 66.7%; Pred. No. 45;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPKWR 6
 I I I
 I I I
 Db 6 FSPLMR 11

RESULT 5
 HSPM_BRAJA STANDARD: PRT: 151 AA.
 AC 086110;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE SMALL HEAT SHOCK PROTEIN HSPH.
 GN HSPH.
 OS Bradyrhizobium japonicum.
 OC Bacteria: Proteobacteria: alpha subdivision: Rhizobiaceae group;
 CC Bradyrhizobium group; Bradyrhizobium.
 OX NCBI_TaxID=375;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Muenchbach M., Nocker A., Narberhaus F.:
 RT Occurrence of a superfamily of small heat shock proteins in
 RT Bradyrhizobium japonicum and other Rhizobium species: a plant-like
 RT phenomenon.
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
 CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AJ010144; CAA09014.1;
 DR InterPro: IPR002068; Crystal11n_HSP20.
 DR Pfam: PF00011; HSP20: 1.
 DR PROSITE: PS01031; HSP20: 1.
 KM Heat shock: Multigene family.
 SQ SEQUENCE 151 AA: 17098 MW: DFE9809310AF6A28 CRC64:

Query Match 90.3%; Score 28; DB 1; Length 151;
 Best Local Similarity 66.7%; Pred. No. 45;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPKWR 6
 I I I
 I I I
 Db 6 FSPLMR 11

RESULT 6
 HSPB_BRAJA STANDARD: PRT: 153 AA.
 AC P70918;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE SMALL HEAT SHOCK PROTEIN HSPB.
 GN HSPB.
 OS Bradyrhizobium japonicum.
 CC Bacteria: Proteobacteria: alpha subdivision: Rhizobiaceae group;

OC Bradyrhizobium group; Bradyrhizobium.
 OX NCBI_TaxID=375;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Netherhaus F., Weiglhofer W., Fischer H.M., Hennecke H.:
 RT "The Bradyrhizobium japonicum rpoH gene encoding a sigma 32-like
 RT protein is part of a unique heat shock gene cluster together with
 RT groESL1 and three small heat shock genes."
 KL J. Bacteriol. 178:5337-5346(1996).
 CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
 CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U55047; AAC44756.1;
 DR InterPro: IPR002068; Crystal11n_HSP20.
 DR Pfam: PF00011; HSP20: 1.
 DR PROSITE: PS01031; HSP20: 1.
 KM Heat shock: Multigene family.
 SQ SEQUENCE 153 AA: 17170 MW: E2E7E24BA69BFD38 CRC64:

Query Match 90.3%; Score 28; DB 1; Length 153;
 Best Local Similarity 66.7%; Pred. No. 45;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPKWR 6
 I I I
 I I I
 Db 7 FAPLMR 12

RESULT 7
 VENV_BEV STANDARD: PRT: 233 AA.
 AC P27904;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-AUG-1992 (Rel. 23, Last annotation update)
 DE ENVELOPE PROTEIN.
 GN E.
 OS Berne virus (BEV).
 OC Viruses: ssRNA positive-strand viruses, no DNA stage: Nidovirales;
 CC Coronaviridae; Torovirus.
 OX NCBI_TaxID=11156;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN-ISOLATE P138/772;
 RT MEDLINE=91220715; PubMed=2024492;
 RA den Boon J.A., Snijder E.J., Locker J.K., Horzinek M.C.,
 RA Rottier P.J.M.:
 RT "Another triple-spanning envelope protein among intracellularly
 RT budding RNA viruses: the torovirus E protein."
 RL Virology 182:655-663(1991).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL: X52505; CAA36747.1;
 DR PIR: A39989; VMWJBV.
 DR PIR: S15570; S15570.
 KM Envelope protein.

SQ SEQUENCE 233 AA: 26548 MW: A07A34DC539104BE CRC64:

Query Match 90.3%: Score 28; DB 1; Length 233;
Best Local Similarity 66.7%: Pred. No. 67;
Matches 4: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPYMR 6
DB 60 FSPYMR 65

RESULT 8
DPSD_CAEEL STANDARD: PRT: 333 AA.

AD 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE PHOSPHATIDYLSETRINE DECARBOXYLASE PROENZYME (EC 4.1.1.65).
CN B0361.5
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodermidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Du 2.1;
RL Submitted (FEB-1994) to the EMBL/Genbank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYL-L-SERINE -
CC PHOSPHATIDYLETHANOLAMINE + CO(2).
CC -1- COFACTOR: REQUIRES A PYRUVYL GROUP FOR ITS ACTIVITY
CC (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U00031; AAK1886.1;
DR Wormpep: B0361.5; CE00834.
DR InterPro: IPR003817; PS_Deacetylase.
DR Pfam: PF02666; PS_Deacetylase; 1.
KW Hypothetical protein; Phospholipid biosynthesis; Lyase; Decarboxylase;
KW Pyruvate; Zymogen.
FT CHAIN 1 299 PHOSPHATIDYLSETRINE DECARBOXYLASE BETA
FT CHAIN (BY SIMILARITY).
FT CHAIN 300 333 PHOSPHATIDYLSETRINE DECARBOXYLASE ALPHA
FT SITE 299 300 CLEAVAGE (NONHYDROLYTICAL)
FT (BY SIMILARITY).
FT MOD_RES 300 300 CONVERTED TO A PYRUVYL GROUP
FT (BY SIMILARITY).
SQ SEQUENCE 333 AA: 37596 MW: 18CF04F2F31E1F34 CRC64:

Query Match 90.3%: Score 28; DB 1; Length 333;
Best Local Similarity 66.7%: Pred. No. 93;
Matches 4: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPYMR 6
DB 36 FXPYMR 41

RESULT 9
CLPK_MYCLE STANDARD: PRT: 508 AA.
ID CLPK_MYCLE
AC O9CB81;

DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GLYCEROL KINASE (EC 2.7.1.30) (ATP:GLYCEROL 3-PHOSPHOTRANSFERASE)
DE (GLYCEROKINASE) (CK).
CN GLPK OR ML2314.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Fellwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagsels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajadream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skellern J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Bartell B.G.;
RT "Massive gene decay in the leprosy bacillus";
RL Nature 409:1007-1011(2001).
CC -1- FUNCTION: KEY ENZYME IN THE REGULATION OF GLYCEROL UPTAKE AND
CC METABOLISM.

CC -1- CATALYTIC ACTIVITY: ATP + GLYCEROL -> ADP + GLYCEROL 3-PHOSPHATE.
CC -1- PATHWAY: RATE-LIMITING STEP IN GLYCEROL UTILIZATION.
CC -1- SIMILARITY: BELONGS TO THE FUCOKINASE / GLUCOKINASE /
CC GLYCEROKINASE / XYLOKINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AL583925; CAC31830.1;
DR Leproma: ML2314;
DR InterPro: IPR000577; FGGY_kin.
DR Pfam: PF00370; FGGY_1.
DR PROSITE: PS00933; FGGY_KINASES_1;
DR PROSITE: PS00445; FGGY_KINASES_2;
KW Glycerol metabolism; Transferase; Kinase; ATP-binding;
KW Complete proteome.
FT NE_BIND 157 169 ATP (PROBABLE).
SQ SEQUENCE 508 AA: 54658 MW: A15379793ECF4039 CRC64:

Query Match 90.3%: Score 28; DB 1; Length 508;
Best Local Similarity 66.7%: Pred. No. 1,4e+02;
Matches 4: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPYMR 6
DB 360 FSPYMR 365

RESULT 10
CLPK_MYCTU STANDARD: PRT: 517 AA.
AC O69664;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GLYCEROL KINASE (EC 2.7.1.30) (ATP:GLYCEROL 3-PHOSPHOTRANSFERASE)
DE (GLYCEROKINASE) (CK).
CN GLPK OR RV3696C OR MT3798 OR MTV025.044C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales: Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE=98295987; Pubmed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeyer K., Cas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver K., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton J., Squares R., Squares S.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Hatt D., Hickey L.,
 RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Ketchum A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: KEY ENZYME IN THE REGULATION OF GLYCEROL UPTAKE AND
 CC METABOLISM.
 CC -1- CATALYTIC ACTIVITY: ATP + GLYCEROL = ADP + GLYCEROL 3-PHOSPHATE.
 CC -1- PATHWAY: RATE-LIMITING STEP IN GLYCEROL UTILIZATION.
 CC -1- SIMILARITY: BELONGS TO THE FUCOKINASE / GLUCOKININASE /
 CC GLYCEROKININASE / XYLUOKININASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AL022121; CAA18018.1; -;
 DR EMBL: AE002177; AAK48165.1; -;
 DR HSSP: P08859; IGLB.
 DR TIGR: MT3798; -;
 DR Tuberculist: RV3696C; -;
 DR InterPro: IPR000577; FGGY_Kin.
 DR Pfam: PF00370; FGGY_1.
 DR PROSITE: PS00445; FGGY_KINASES_2; 1.
 DR PROSITE: PS00933; FGGY_KINASES_1; 1.
 DR GlycoLip: G00001; Transferrase; Kinase; ATP-binding;
 KW Complete proteome.
 FT NP_BIND 165 177 ATP (PROBABLE).
 FT SEQUENCE 517 AA; 55859 MW; 7E3F30DCFA1B63E CRC64;
 SQ

IT 01-AUG-1990 (Rel. 15, Last annotation update)
 DE 69.7 KDA PROTEIN (ORF 2).
 CS Potato leafroll virus (strain Wageningen) (PURV).
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
 CC Polerovirus.
 OX NCBI_TaxID=12048;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89171329; Pubmed=2466700;
 RA van der Wilk F., Huismans M.J., Cornelissen B.J.C., Huitinga H.,
 RA Goldbach R.W.;
 RT "Nucleotide sequence and organization of potato leafroll virus
 RT genomic RNA.";
 RL FEBS Lett. 245:51-56(1989).
 CC -1- SIMILARITY: 54% SIMILARITY TO BEET WESTERN YELLOW VIRUS ORF2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Y07496; CAA68795.1; -;
 DR PIR: S03547; S03547.
 DR InterPro: IPR000382; Luteo_ORF2.
 DR Pfam: PF02122; Luteo_ORF2; 1.
 DR PRINTS: PR00913; LVIRUSORF2.
 SQ SEQUENCE 639 AA; 69676 MW; 92E1473FE3FEF148 CRC64;
 RESULT 12
 YAF7_CAEEL
 ID YAF7_CAEEL STANDARD: PRT: 199 AA.
 AC P52883;
 IRT 01-OCT-1996 (Rel. 34, Created)
 IRT 01-OCT-1996 (Rel. 34, Last sequence update)
 IRT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE HYPOHETICAL 23.2 KDA PROTEIN P46C5.7 IN CHROMOSOME II.
 UN P46C5.7.
 CS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 CC Rhabditidae; Pelodieridae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Thomas R.;
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z54281; CAA91049.1; -;
 DR WormPep: P46C5.7; CE03348.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 35 55 POTENTIAL.

FT TRANSMEM 57 77 POTENTIAL.
 FT TRANSMEM 94 114 POTENTIAL.
 FT TRANSMEM 131 151 POTENTIAL.
 SO SEQUENCE 199 AA: 23247 MW: AC09222A692E3638 CRC64:

Query Match 87.1%; Score 27; DB 1; Length 199;
 Best Local Similarity 66.7%; Pred. No. 90;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXWR 6
 1 1 1 1
 DB 22 FMPFWR-27

RESULT 13
 HO_FUGRU STANDARD: PRT: 277 AA.
 ID HO_FUGRU
 AC 073688;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE HEME OXYGENASE (EC 1.14.99.3) (HO).
 OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontidae; Takifugu.
 OX NCBI_TaxID=31033;
 RN 11
 RP SEQUENCE FROM N.A.
 RX MEDLINE-98163746; PubMed-9503016;
 RA Cottens B., Gilbert J.G.R., Barton L.M., Aparicio S., Hawker K.,
 RA Mistry S., Vaudin M., King A., Bentley D., Elgar G., Green A.R.,
 RT The pufferfish SLF-1 gene, a new member of the SCL/TAL-1 family of
 RT Transcription factors."
 RL Genomics 48:52-62(1998).
 CC -1- FUNCTION: HEME OXYGENASE CLEAVES THE HEME RING AT THE ALPHA
 CC METHENE BRIDGE TO FORM BILIVERDIN. BILIVERDIN IS SUBSEQUENTLY
 CC CONVERTED TO BILIRUBIN BY BILIVERDIN REDUCTASE. UNDER
 CC PHYSIOLOGICAL CONDITIONS, THE ACTIVITY OF HEME OXYGENASE IS
 CC HIGHEST IN THE SPLEEN, WHERE SENESCENT ERYTHROCYTES ARE
 CC SEQUESTERED AND DESTROYED (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: HEME + 3 AH(2) + O(2) -> BILIVERDIN + FE(2+) +
 CC CO + 3 A + 3 H(2)O.
 CC -1- SUBCELLULAR LOCATION: MICROSOMAL (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE HEME OXYGENASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: AF022814; AAC41263.1;
 DR InterPro: IPR002051; Heme_oxygenase.
 DR Pfam: PF01126; Heme_oxygenase; 1.
 DR PRINTS: PRO0088; HAEMOXGNASE.
 DR PROSITE: PS00593; HEME_OXYGENASE; 1.
 KM Heme: Oxidoreductase; Microsome.
 FT BINDING 29 29 PROXIMAL HEME LIGAND (BY SIMILARITY).
 FT SEQUENCE 277 AA: 31211 MW: 7783584699963777 CRC64;

Query Match 87.1%; Score 27; DB 1; Length 277;
 Best Local Similarity 66.7%; Pred. No. 1,2e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXWR 6
 1 1 1 1
 DB 101 FCPDWR 106

RESULT 14
 BCS1_YEAST STANDARD: PRT: 456 AA.
 ID BCS1_YEAST
 AC P32839; Q06404;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE BCS1 PROTEIN.
 GN BCS1 OR YDR375C OR D9481.17.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN 11
 RP SEQUENCE FROM N.A.
 RX MEDLINE-93010976; PubMed-1327750;
 RA Nobrega F.G., Nobrega M.P., Tzagoloff A.;
 RT "BCS1, a novel gene required for the expression of functional Rieske
 RT iron-sulfur protein in Saccharomyces cerevisiae."
 RL EMBO J. 11:3821-3829(1992).
 RN 12
 RP SEQUENCE FROM N.A.
 RX STRAIN=5288C / AB972;
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
 RA Favell A., Fulton L., Gattung S., Greco T., Kirsten J.,
 RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
 RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
 RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
 RA Rifken L., Riles L., Tatch A., Trevasakis E., Vignati D.,
 RA Wilcox L., Wohlman P., Vaudin M., Wilson R., Waterston R.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 RN 13
 RP TOPOLOGY.
 RX MEDLINE-96174808; PubMed-8599931;
 RA Foelisch H., Guillard B., Neupert W., Stuart R.A.;
 RT "Internal targeting signal of the BCS1 protein: a novel mechanism of
 RT import into mitochondria."
 RL EMBO J. 15:479-487(1996).
 CC -1- FUNCTION: ESSENTIAL FOR THE EXPRESSION OF THE RIESKE IRON-SULFUR
 CC PROTEIN.
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. MITOCHONDRIAL.
 CC -1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: S47190; AAC09007.1;
 DR EMBL: U28373; AAB64811.1;
 DR PIR: S28411; S28411.
 DR SGD: S0002783; BCS1.
 DR InterPro: IPR003593; AAA.
 DR InterPro: IPR001939; AAA_subfam.
 DR Pfam: PF00004; AAA; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00674; AAA; 1.
 KW ATP-binding; Mitochondrion; Transmembrane.
 FT DOMAIN 1 44 MITOCHONDRIAL INTERMEMBRANE (PROBABLE).
 FT TRANSMEM 45 68 PROBABLE.
 FT DOMAIN 69 456 MITOCHONDRIAL MATRIX (PROBABLE).
 FT SITE 69 83 MITOCHONDRIAL TARGETING SIGNAL.
 FT NE_BIND 267 274 ATP (POTENTIAL).
 FT CONFLICT 119 119 N -> F (IN REF. 1).
 FT CONFLICT 267 267 G -> S (IN REF. 1).
 FT CONFLICT 267 267 G -> S (IN REF. 1).
 FT CONFLICT 349 349 L -> Q (IN REF. 1).
 FT CONFLICT 351 351 G -> S (IN REF. 1).
 FT SEQUENCE 456 AA: 51107 MW: 17750D26B64ED04 CRC64;

Query Match 87.1%; Score 27; DB 1; Length 456;

Best Local Similarity 66.7%; Pred. No. 1.9e+02; Mismatches 2; Indels 0; Gaps 0;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EXPXMR 6
1 1 1 1
Db 209 FCPGMR 214

RESULT 15

RBL2_RHOSH STANDARD; PRT: 459 AA.

AC 01-DEC-1992 (Rel. 24, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DE 15-DEC-1998 (Rel. 37, Last annotation update)

DE RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN 2 (EC 4.1.1.39)

DE (RUBISCO LARGE SUBUNIT).

CN CBBL2 OR RBL.

OS Rhodospirillum rubrum (Rhodospirillum rubrum).

OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum group;

OC Rhodospirillum.

OX NCBI_TaxID=1063;

RN 11

RP SEQUENCE FROM N.A.

RA Wagner S.J., Stevens S.E. Jr., Nixon B.T., Lambert D.H.,

RT Quilley R.G. Jr., Tabita F.R.;

RT "Nucleotide and deduced amino acid sequence of the Rhodospirillum

RT spheeroides gene encoding form II ribulose-1,5-bisphosphate

RT carboxylase/oxygenase and comparison with other deduced forms I and II

RT sequences."

RL FEWS Microbiol. Lett. 55:217-222(1988).

RN 12

RP SEQUENCE OF 1-36 FROM N.A.

RA MEDLINE=92041881; PubMed=1939098;

RT Chen J.-H., Gibson J.L., McCue L.A., Tabita F.R.;

RT "Identification, expression, and deduced primary structure of

RT transketolase and other enzymes encoded within the form II CO2

RT fixation operon of Rhodospirillum rubrum."

RL J. Biol. Chem. 266:20447-20452(1991).

CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF

CC D-RIBULOSE 1,5-BISPHOSPHATE. THE PRIMARY EVENT IN PHOTOSYNTHETIC

CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF

CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH

CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME

CC ACTIVE SITE.

CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) =

CC 2 3-PHOSPHO-D-GLYCERATE.

CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) =

CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.

CC -1- SUBUNIT: IN CONTRAST TO FORM I GENES, THE FORM II RUBISCO ARE

CC COMPOSED OF ONLY LARGE SUBUNITS.

CC -1- MISCELLANEOUS: THIS PROTEIN IS ENCODED WITHIN THE FORM II

CC RIBULOSE-BISPHOSPHATE CARBOXYLASE OPERON.

CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: M68914; AAA26158.1; ALT_SEQ.

CC DR PIK: E41080; E41080.

CC DR HSSP: P04718; 1RBA.

CC DR InterPro: IPR000685; Rubisco_large.

CC DR Pfam: PF00016; Rubisco_large; 1.

CC DR PROSITE: PS00157; RUBISCO_LARGE; 1.

CC KM Photosynthesis; Carbon dioxide fixation; Photorespiration;

KM Lyase: Oxidoreductase; Monooxygenase; Multigene family.

ACT SITE 191 BINDING OF CO(2) ACTIVATES THE ENZYME.

SEQUENCE 459 AA; 50519 MW; 299ABAA836BD683E CRC64;

Query Match 87.1%; Score 27; DB 1; Length 459;

Best Local Similarity 66.7%; Pred. No. 1.9e+02; Mismatches 2; Indels 0; Gaps 0;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EXPXMR 6
1 1 1 1
Db 447 FYPGMR 452

Search completed: February 27, 2002, 11:42:47

Job time: 546 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:50:11 ; Search time 281.76 Seconds
(without alignments)
3.115 Million cell updates/sec

Title: US-09-446-109A-12
Perfect score: 31
Sequence: 1 FXPXWR 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 14627329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP archaea: *
2: SP bacteria: *
3: SP fungi: *
4: SP human: *
5: SP invertebrate: *
6: SP mammal: *
7: SP mnc: *
8: SP organelle: *
9: SP phage: *
10: SP plant: *
11: SP rodent: *
12: SP virus: *
13: SP vertebrate: *
14: SP unclassified: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	93.5	296	2 033353	033353 mycobacteri
2	29	93.5	354	3 060121	060121 schizosacch
3	29	93.5	563	2 P72421	P72421 salmonella
4	29	93.5	572	12 09J8B6	09J8B6 spodoptera
5	28	90.3	135	2 09P873	09P873 xylella fas
6	28	90.3	141	2 09WKK5	09WKK5 streptococc
7	28	90.3	143	5 096402	096402 plasmidium
8	28	90.3	160	4 095530	095530 homo sapien
9	28	90.3	160	4 09BQ29	09BQ29 homo sapien
10	28	90.3	211	2 033867	033867 bacillus pu
11	28	90.3	218	4 095726	095726 homo sapien
12	28	90.3	221	10 09XJ12	09XJ12 oryza sativ
13	28	90.3	236	2 005115	005115 methyllobact
14	28	90.3	241	5 09VLD3	09VLD3 drosophila
15	28	90.3	250	2 09EUM2	09EUM2 corynebacte
16	28	90.3	254	2 049185	049185 mycobacteri
17	28	90.3	254	2 09FIE6	09FIE6 acetobacter
18	28	90.3	264	2 09FDM9	09FDM9 salmonella
19	28	90.3	272	11 09DA91	09DA91 mus musculu

20	28	90.3	297	10 0920C7	0920C7 arabidopsis
21	28	90.3	311	5 044789	044789 caenorhabdi
22	28	90.3	317	2 09L4F0	09L4F0 bacillus ce
23	28	90.3	325	2 09L4E9	09L4E9 bacillus ce
24	28	90.3	337	1 09HP24	09HP24 halobacteri
25	28	90.3	357	5 026862	026862 trypanosoma
26	28	90.3	389	2 0910K8	0910K8 pseudomonas
27	28	90.3	407	2 09A459	09A459 caulobacter
28	28	90.3	410	2 09A108	09A108 pinus strob
29	28	90.3	475	1 026479	026479 methanobact
30	28	90.3	483	2 09A8C6	09A8C6 caulobacter
31	28	90.3	488	10 041017	041017 pinus strob
32	28	90.3	490	2 09A313	09A313 caulobacter
33	28	90.3	507	2 09RJM2	09RJM2 streptomyc
34	28	90.3	512	2 09ADA7	09ADA7 streptomyc
35	28	90.3	542	10 09SHX0	09SHX0 arabidopsis
36	28	90.3	552	2 006335	006335 mycobacteri
37	28	90.3	599	4 09H5P1	09H5P1 homo sapien
38	28	90.3	624	4 09C0D7	09C0D7 homo sapien
39	28	90.3	633	4 09Y4W3	09Y4W3 homo sapien
40	28	90.3	635	2 09Z7B3	09Z7B3 chlamydia p
41	28	90.3	635	2 09JSA3	09JSA3 chlamydia p
42	28	90.3	639	12 084836	084836 potato leaf
43	28	90.3	703	10 09M1P7	09M1P7 arabidopsis
44	28	90.3	781	4 095560	095560 homo sapien
45	28	90.3	791	4 09C0B6	09C0B6 homo sapien

ALIGNMENTS

RESULT 1
033353 PRELIMINARY: PRT; 296 AA.
ID 033353
AC 033353;
DT 01-JAN-1998 (TEMBREL. 05, Created)
DT 01-JUN-1998 (TEMBREL. 06, last sequence update)
DT 01-MAY-2000 (TEMBREL. 13, last annotation update)
DE HYPOTHEICAL 32.0 KDA PROTEIN (ORF2).
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBL_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-IS1547;
RX MEDLINE=9121044; PubMed=9003621;
FA Fang Z., Forbes K.J.;
RT "A Mycobacterium tuberculosis IS6110 preferential locus (lpl) for
RT insertion into the genome."
KL J. Clin. Microbiol. 35:479-481(1997).
DR EMBL: Y13470; CAA73869.1;
KW Transposable element; Hypothetical protein.
SQ SEQUENCE 296 AA; 31996 MW; 3CDACBC956653F0 CRC64;

Query Match 93.5%; Score 29; DB 2; Length 296;
Best local Similarity 66.7%; Pred. No. 3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Caps 0;

QY 1 FXPXWR 6
DB 263 FSPSWR 268

RESULT 2			
ID	060121	PRELIMINARY:	PRT: 354 AA.
AC	060121:		
DT	01-AUG-1998 (TREMBLrel. 07, Created)		
DT	01-AUG-1998 (TREMBLrel. 07, Last sequence update)		
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)		
DE	HYPOTHEETICAL 39.3 KDA PROTEIN C16G5.07C IN CHROMOSOME II.		
CN	SPBIC16G5.07C.		
OS	Schizosaccharomyces pombe (Fission yeast).		
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;		
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;		
OC	Schizosaccharomycetes.		
OX	NCBI_TaxID=4896;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=972;		
RA	Lyne M., Wood V., Rajandream M.A., Barrell B.G., Beck A.,		
RA	Reinhardt R.;		
RL	Submitted (MAR-1998) to the EMBL/GenBank/DBJ databses.		
CC	-1- SIMILARITY: BELONGS TO THE BAND 7 / MEC-2 FAMILY.		
CC	EMBL; AL023554; CAA19027.1; -		
DR	InterPro: IPR001107; Band.7.		
DR	InterPro: IPR001972; Stomatin.		
DR	Pfam: PF01145; Band.7; 1.		
DR	SMART: SM00244; PHB; 1.		
DR	PROSITE: PS01270; BAND_7; FALSE_NEG.		
KW	Hypothetical protein.		
SO	SEQUENCE 354 AA: 39274 MW: 66D09E6A12BDC030 CRC64:		

Query Match	93.5%	Score 29	DB 3	Length 354
Best Local Similarity	66.7%	Pred No.	3.5e+02	
Matches 4	Conservative	0	Mismatches 2	Indels 0
			Gaps	0

QY	1 EXPXWR-6
	1 1 1 1
Db	40 ETPYWR 45

RESULT	3
P72421	
ID	P72421
AC	P72421
DT	01-FEB-1997 (TREMBLrel_02, Created)
DT	01-FEB-1997 (TREMBLrel_02, Last sequence update)
DT	01-JUN-2001 (TREMBLrel_17, Last annotation update)
DE	(SEROTYPE TYPHIMURIUM).
CN	MKFB.
OS	Salmonella typhimurium.
OC	Plasmid.
CC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC	Salmonella.
OX	NCB1_TaxID=602;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-SEROTYPE TYPHIMURIUM, C5;
RA	MEDLINE:90161559; Pubmed:2696057.
RT	Note1 F., Pisanò M.R., Nicoll J.J., Popoff M.Y.;
RT	"Nucleotide sequence of the plasmid-borne virulence gene mktB from
RL	Salmonella typhimurium.";
Res.	Microbiol. 140:455-457(1989).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN-SEROTYPE TYPHIMURIUM, C5;
RA	Note1 F.;
DL	Submitted (JAN-1992) to the EMBL/GenBank/DDBJ databases.
DR	EMBL: X57096; CA40380.1; .
DR	InterPro: IPR003284; Sal_SpvB.
DR	PRINTS: PR01341; SALSPVBPROT.
KW	plasmid.
SQ	SEQUENCE 563 AA: 62586 MW: DA3F093ADB8241CDA CRC64:

Query Match	93.5%	Score 29	DB 2	Length 563
Best Local Similarity	66.7%	Pred. No.	5.5e+02	
Matches	4	Conservative	0	Mismatches 2
				Indels 0
				Gaps 0
Qy	1	EXPXMR	6	
+	1	1	1	
DB	50	FAPXMR	55	

QY	1	EXPXWR	6
∴		1 1 1 1	
Db	50	FAPAWR	55

RESULT 4
COTRANS

RESULT	4			
ID	C9J8B6	PRELIMINARY:	PRT:	572 AA.
AC	Q9J8B6.			
DT	01-OCT-2000 (TrEMBLrel. 15, Created)			
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)			
DT	01-JUN-2001 (TrEMBLrel. 15, last annotation update)			
DE	ORF19 CHITINASE.			
OS	Spodoptera exigua nucleopolyhedrovirus.			
OC	Viruses; dsDNA viruses, no RNA stage; Baculoviridae;			
OX	NCBI_TaxID=10454;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20036646; Pubmed=10567663;			
RA	Itsek W.F., van Strien E.A., Heldens J.G., Broer R., Zuidema D.			
RA	Goldbach R.W., Vlak J.M.;			
RT	"Sequence and organization of the spodoptera exigua multicapsid			
KL	nucleopolyhedrovirus genome.";			
KL	J. Gen. Virol. 80:3289-3304(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Itsek W.F.J., van Strien E.A., Heldens J.G.M., Broer R., Zuidema			
RA	Goldbach R.W., Vlak J.M.;			
RL	Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.			
DR	EMBL; AF169823; AF35349.1; -			
DR	InterPro; IPR000886; ER_target.			
DR	InterPro; IPR001223; Glyco_hydro_18.			
DR	InterPro; IPR021173; Pfam.			
DR	Pfam; PF00704; Glyco_hydro_18; 1.			
DR	PROSITE; PS00014; ER_TARGET; UNKNOWN_1			
DR	PROSITE; PS00583; PFRK_KINASES_1; UNKNOWN_1			
SO	SEQUENCE 572 AA: 63508 MW: A678AAD14021A014 CRC64:			

Query Match	93.5%	Score 29	DB 12	length 572
Best Local	Similarity 66.7%	Pred NO. 5.5e+02		
Matches 4	Conservative 0	Mismatches 2	Indels 0	Gaps 0

OY	1	FXPRWR	6
Db	401	FAPTRW	400

RESULT	5	
09PB73		
ID	Q9PB73	PRELIMINARY; PRT; 135 AA.
AC	Q9PB73;	
DT	01-OCT-2000 (TREMBLrel, 15, Created)	
DT	01-OCT-2000 (TREMBLrel, 15, Last sequence update)	
LT	01-MAR-2001 (TREMBLrel, 16, Last annotation update)	
DE	HYDROTHERMAL PROTEIN XE2271.	
CN	XP2271.	
OS	Xyella fastidiosa.	
OC	Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;	
CC	Xyella.	
OX	NCBI_TaxID=2371;	
SN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=9A5C;	
RX	MEDLINE=20365717; PubMed=1091047.	
RA	Simpson A.J.G., Reinbach F.C., Artida P., Abreu F.A., Agencio M.,	
RA	Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,	

RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carreir H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofolini M., Dias-Neto E., Docena C., El-Dorty H.,
RA Faccinani A.P., Ferrelha A.J.S., Ferrelha V.C.A., Ferro J.A.,
RA Fraga J.S., Frasca S.C., Franco M.C., Frohne M., Furlan L., Gruber A.,
RA Gantier M., Goldman G.H., Goldman M.H.S., Gomes S., Lull, L.R.B.,
RA Ho P.L., Hohseil J.D., Junqueira M.L., Kemper E.L., Kitzljma J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.C.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madella A.M.B.N., Madella H.M.F., Marinho A.V.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Merck C.F.M., Miracca E.C., Miyaki C.Y., Monteloro-Vicorello C.B.,
RA Moon D.H., Nagai M., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A., Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Pais A.,
RA Peixoto B.R., Petella G.A.G., Petella H.A., Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E., Jr., de Sa R.G., Sattelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A., Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuchako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zaitz M., Meidanis J., Setubal J.C.,
RT "The genome sequence of the plant pathogen *Xylella fastidiosa*,"
RL Nature 406:151-159(2000).
DR EMBL: AE004039; AAF85070.1; -
KW Hypothetical protein; Complete proteome
SQ SEQUENCE 135 AA: 15203 MW: 99E7F0EA22205244 CRC64;

Query Match	90.3%	Score 28	DB 2	Length 135
Best Local Similarity	66.7%	Pred. No. 2	2e+02	
Matches	4	Conservative	0	Mismatches 2
		Indels	0	Gaps 0

OY	1	EXPXMR	6
	1	1	1
	1	1	1
Db	79	FVPAWR	84

RESULT	6
O9MWKS	
ID	O9MWKS
AC	O9MWKS
DT	01-NOV-1999 (TREMblrel. 12, Created)
DT	01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT	01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE	HYPOTHEtical 16.0 KDA PROTEIN.
OS	Synechococcus sp. (Strain PCC 7002) (Agmenellum quadruplicatum).
OC	Bacteria: Cyanobacteria: Chroococcales: Synechococcus.
OX	NCRL_TaxID=13049;
RN	11
RP	SEQUENCE FROM N. A.
RC	STRAIN-PCC7002, PR6000.
RA	Schlink K., Kowalik K.V., Bryant D.A.;
RL	Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
DR	EMBL: AF136577; AAD26588.1; -
DR	HSSP: P00210; 1FXR.
DR	InterPro: IPR001080; 3fe4s.ferrdxn.
DR	PRINTS: PR00352; 3FE4SFRRDXIN.
KW	HYPOthetical protein.
QW	SEQUENCE 141 AA; 16034 MW; 85872P255B76FCE2 CRC64;

Query Match	90.3%	Score 28	DB 2	Length 141
Best Local Similarity	66.7%	Pred. No. 2.3e+02		
Matches	4	Conservative	0	Mismatches 2
				Indels 0
				Gaps 0
Oy	1	EXPXMR	6	
	1	1	1	
	1	1	1	
Db	122	FSPGMR	127	

	RESULT	7			
C96402					
ID	096402	PRELIMINARY:	PRT:	143 AA.	
AC	096402:				
DT	01-MAY-1999 (TREMBLrel)	10.	Created		
DT	01-MAY-1999 (TREMBLrel)	10.	Last sequence update)		
DT	01-MAY-1999 (TREMBLrel)	10.	Last annotation update)		
DE	ERYTHROCYTE MEMBRANE PROTEIN 1 (FRAGMENT).				
GN	PCR3SL.ZVAR6.				
OS	Plasmodium falciparum.				
OC	Eukaryota; Alveolata;	Apicomplexa;	Haemosporida;	Plasmodium	
OX	NCBI_TaxID=5833:				
RN	(1)				
R2	SEQUENCE FROM N.A.				
RA	Chen O., Wahlgren M.				
RL	Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.				
RR	EMBL: AF039278; ADDO2161.1); -.				
FT	NON_TER	143	143		
FT	NON_TER	1	1		
SQ	SEQUENCE	143 AA:	16818 MW:	A8641F68BE09B3CE CRC64:	

Query Match	90.3%	Score 28	DB 5	Length 143
Best Local Similarity	66.7%	Pred. No. 2.4e+02		
Matches	4	Conservative	2	Mismatches 0
			Indels	0
			Gaps	0
Cy	1 EXPXMR 6			
	1 1 1 1			
Db	2 ENPTWR 7			

RESULT	8	
Q95530		
ID	Q95530	PRELIMINARY: PRF: 160 AA.
NC	Q95530:	
DT	01-MAY-1999. (TREMBL)	10, Created)
DT	01-MAY-1999. (TREMBL)	10, Last sequence update)
DE	DJ281H8.2 (PUTATIVE NOVEL PROTEIN SIMILAR TO KIA0312) AND WORM	
GN	DJ281H8.2.	
OS	Homo sapiens (human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Mashreghi-Mohammadi M.;	
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.	
LR	EMBL: AL031133; CAA20017.1; -	
FT	NON_TER	1
FT	NON_TER	160
ST	SEQUENCE	160 AA; 18983 MW; EC3F622FFC4E4F93 CMC64;

```

Query Match Similarity      90.3%  Score 28: DB 4: Length 160;
Best Local Similarity      66.7%  Pred. No. 2.6e+02;
Matches      4:  Conservative      0:  Mismatches      2:  Indels      0:  Gaps      0:
DY      1  EXPXMR  6
      1  1  1  1
DB      30  FVPSMR  35

RESULT      9
Q9BOZ9
ED  Q9BOZ9:      PRELIMINARY:      PRT:      160 AA.
SC  Q9BOZ9:
DT      01-JUN-2001 (TREMBLrel. 17, Created)
DT      01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT      01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE      D3423R2.1 (NOVEL PROTEIN SIMILAR TO KIAA0123, KIAA0615 AND C.ELEGANS
DE      C30F12.1) (FRAGMENT).

```

GN DJ423B22.1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Pearce A.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL034379; CAC36083.1;
 FT NON_TER 1 160
 FT SEQUENCE 160 AA; 18953 MW; 53EEA2B52DAED6C0 CRC64;
 SQ

Query Match 90.3%; Score 28; DB 4; Length 160;
 Best Local Similarity 66.7%; Pred. No. 2.6e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPMR 6
 DB 30 FVPSMR 35

RESULT 10
 ID 031867 PRELIMINARY; PRT: 211 AA.
 AC 031867;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE SIMILAR TO ANSR GENBANK ACCESSION NUMBER Q07683.
 OS Bacillus pumilus (Bacillus mesentericus).
 OC Plasmid pSH1452.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1408;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-SH1451;
 RA Thomas C.W., Hasnain S.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U53767; AAB71491.1;
 DR InterPro: IPR001387; HTH_3.
 DR Pfam: PF01381; HTH_3; 1
 DR SMART: SM00530; HTH_XRE; 1.
 KW Plasmid.
 SQ SEQUENCE 211 AA; 24620 MW; 49BD5954BD35E83A CRC64;

Query Match 90.3%; Score 28; DB 2; Length 211;
 Best Local Similarity 66.7%; Pred. No. 3.4e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPMR 6
 DB 199 F1PTMR 204

RESULT 11
 ID 095726 PRELIMINARY; PRT: 218 AA.
 AC 095726;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
 DE HYPOTHETICAL 25.5 KDA PROTEIN (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.

KC TISSUE-BRAIN;
 RA Park S.H., Kim C., Kang Y.H., Chung H.S., Kim H.;
 KT *Expression of novel gene mapped on chromosome 7q22.*;
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U67037; AAD09521.1;
 KW Hypothetical protein.
 FT NON_TER 1 218
 FT SEQUENCE 218 AA; 25458 MW; C0896F5EDAC5627F CRC64;
 SQ

Query Match 90.3%; Score 28; DB 4; Length 218;
 Best Local Similarity 66.7%; Pred. No. 3.5e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPMR 6
 DB 202 FVPSMR 207

RESULT 12
 ID 09XJ12 PRELIMINARY; PRT: 221 AA.
 AC 09XJ12;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
 DE HYPOTHETICAL PROTEIN.
 OS Oryza sativa (rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriocaridaceae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. NIPPONBARE;
 RA Sasaki T., Nagamura Y., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC
 clone: p0680A03."
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB023482; BAA78741.1;
 SQ SEQUENCE 221 AA; 24502 MW; DC24FF370660DC15 CRC64;

Query Match 90.3%; Score 28; DB 10; Length 221;
 Best Local Similarity 66.7%; Pred. No. 3.5e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPMR 6
 DB 203 FVPSMR 208

RESULT 13
 ID 005115 PRELIMINARY; PRT: 236 AA.
 AC 005115;
 DT 01-JUL-1997 (TREMBlrel. 04, Created)
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE PUTATIVE GLYCERATE KINASE AND PYRUVATE KINASE (PYKA) GENES, COMPLETE
 DE CDS (PYKA).
 OS Methylobacterium extorquens.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Methylobacterium group; Methylobacterium.
 OX NCBI_TaxID=408;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-AM1;
 RX MEDLINE-97386438; PubMed-9244287;
 RA Chistoserdova L., Lidstrom M.E.;
 KT "Identification and mutation of a gene required for glycerate kinase
 activity from a facultative methylotroph, Methylobacterium extorquens

RT AML.
 RL J. Bacteriol. 179:4946-4948(1997).
 DR EMBL: U07316; AAB66495.1;
 SO SEQUENCE 236 AA; 26139 MW; 7F48376B148B3676 CRC64;

Query Match
 Best Local Similarity 66.7%; Score 28; DB 2; Length 236;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXMR 6
 1 1 1 1
 Db 135 FTFPMR-140

RESULT 14
 ID 09VLD3 PRELIMINARY; PRT: 241 AA.
 AC 09VLD3:
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE CG3694 PROTEIN.
 GN CG3694.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 OX (1)
 RN SEQUENCE FROM N.A.
 RP STRAIN=BERKELEY;
 RC MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Plamkoc C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Correll J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeguam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matvei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshnell A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K.A., Nusskern D.R., Paclet J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003624; AAF52759.1;
 DR FlyBase: FBgn0032107; CG3694.

SO SEQUENCE 241 AA; 27071 MW; 200925FE35C3E733 CRC64;

Query Match
 Best Local Similarity 66.7%; Score 28; DB 5; Length 241;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXMR 6
 1 1 1 1
 Db 68 FTFPMR 73

RESULT 15
 ID 09EUM2 PRELIMINARY; PRT: 250 AA.
 AC 09EUM2:
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE YCG4L.
 GN YCG4L.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OG Plasmid R-plasmid pCG4.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;
 OC Corynebacterium.
 OC NCBI_TaxID=1718;
 OX (1)
 RN SEQUENCE FROM N.A.
 RP STRAIN=ATCC31830;
 RA Tauch A., Puchler A., Kalinowski J.,
 RT "DNA sequence and genetic organization of the integron-carrying R-
 Subcloned pCG4 of Corynebacterium glutamicum".
 RL EMBL: AF164956; AAG00294.1;
 DR InterPro: IPR002145; CopC_HTH_4.
 DR Pfam: PF01402; HTH_4; 1.
 KW Plasmid.
 SO SEQUENCE 250 AA; 28804 MW; 195C86367B483157 CRC64;

Query Match
 Best Local Similarity 66.7%; Score 28; DB 2; Length 250;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXMR 6
 1 1 1 1
 Db 214 FXPXMR 219

Search completed: February 27, 2002, 11:50:12
 Job time: 991 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:36:05 ; Search time 132.19 Seconds
(without alignments)
1.021 Million cell updates/sec

Title: US-09-446-109A-12
Perfect score: 31
Sequence: 1 FXPXMR 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/laa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/laa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/laa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/laa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/laa/PCrUS_COMB.pep:*
6: /cgn2_6/ptodata/2/laa/backfill1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	90.3	527	3	US-08-907-229-2
2	28	90.3	1536	4	US-09-413-814-10
3	28	90.3	2232	4	US-09-091-219-25
4	28	90.3	2247	4	US-09-091-219-2
5	27	87.1	6	1	US-08-215-137-13
6	27	87.1	456	4	US-09-268-364-21
7	26	83.9	42	2	US-08-766-858A-27
8	26	83.9	254	4	US-09-362-831-3
9	26	83.9	273	2	US-08-997-080-75
10	26	83.9	273	2	US-08-997-362-75
11	26	83.9	273	3	US-08-873-970-75
12	26	83.9	273	3	US-08-095-855-75
13	26	83.9	370	2	US-08-997-080-194
14	26	83.9	370	2	US-08-997-362-194
15	26	83.9	370	4	US-09-095-855-194
16	26	83.9	410	4	US-09-238-480-2
17	26	83.9	604	2	US-08-735-041A-4
18	26	83.9	604	3	US-09-190-476B-4
19	26	83.9	604	3	US-09-190-889A-4
20	26	83.9	604	4	US-09-190-938B-4
21	26	83.9	701	2	US-08-735-041A-6
22	26	83.9	701	3	US-09-190-476B-6
23	26	83.9	701	3	US-09-190-889A-6
24	26	83.9	701	4	US-09-190-938B-6
25	25	80.6	38	1	US-08-176-500-118
26	25	80.6	38	1	US-08-471-052A-118
27	25	80.6	38	1	US-08-189-331-118

28	25	80.6	38	2	US-08-471-939-118	Sequence 118, App
29	25	80.6	38	2	US-08-471-800-118	Sequence 118, App
30	25	80.6	38	2	US-08-471-068-118	Sequence 118, App
31	25	80.6	175	4	US-09-060-726A-2	Sequence 2, Appl
32	25	80.6	631	4	US-08-487-890A-115	Sequence 115, App
33	25	80.6	631	2	US-08-478-435-115	Sequence 115, App
34	25	80.6	631	2	US-08-337-483-115	Sequence 115, App
35	25	80.6	631	2	US-08-478-373-115	Sequence 115, App
36	25	80.6	631	3	US-08-474-671-115	Sequence 115, App
37	25	80.6	631	3	US-08-483-577A-115	Sequence 115, App
38	25	80.6	631	4	US-08-897-438-115	Sequence 115, App
39	25	80.6	743	4	US-09-077-354B-2	Sequence 2, Appl
40	24	77.4	10	3	US-08-159-339A-448	Sequence 48, App
41	24	77.4	15	2	US-08-553-257A-52	Sequence 52, Appl
42	24	77.4	175	1	US-08-414-926A-8	Sequence 8, Appl
43	24	77.4	175	2	US-08-926-922-8	Sequence 8, Appl
44	24	77.4	175	3	US-09-253-682-8	Sequence 8, Appl
45	24	77.4	175	4	US-09-527-657-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1
US-08-907-229-2
Sequence 2, Application US/08907229A
Patent No. 6072048
GENERAL INFORMATION:
APPLICANT: Riley, Lee W.
TITLE OF INVENTION: DNA MOLECULE ENCODING FOR CELLULAR UPTAKE OF
FILE REFERENCE: 19603/1531
CURRENT APPLICATION NUMBER: US/08/907, 229A
CURRENT FILING DATE: 1997-08-06
EARLIER APPLICATION NUMBER: 60/040, 097
EARLIER FILING DATE: 1997-03-10
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 527
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-08-907-229-2

Query Match 90.3%, Score 28; DB 3; Length 527;

Best Local Similarity 66.7%; Pred. No. 2.9e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXMR 6
DB 3 FXPXMR 8

RESULT 2
US-09-413-814-10
Sequence 10, Application US/09413814
Patent No. 6225064
GENERAL INFORMATION:
APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
APPLICANT: Bristol-Myers Squibb, Co.
APPLICANT: Beyer, Stefan
APPLICANT: Bloeker, Helmut
APPLICANT: Brandt, Petra
APPLICANT: Cino, Paul M
APPLICANT: Dougherty, Brian A
APPLICANT: Goldberg, Steven L
APPLICANT: Hoffe, Gerhard
APPLICANT: Mueller, Joachim
APPLICANT: Reichenbach, Hans
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
TITLE OF INVENTION: heteropolypeptide compounds
FILE REFERENCE: PCT/US 99/23535

;; CURRENT APPLICATION NUMBER: US/09/413.814
;; CURRENT FILING DATE: 1999-10-07
;; EARLIER APPLICATION NUMBER: DE 198 46 493.2
;; EARLIER FILING DATE: 1998-10-09
;; NUMBER OF SEQ ID NOS: 107
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 10
;; LENGTH: 1536
;; TYPE: PRT
;; ORGANISM: Sorangium cellulosum
;; US-09-413-814-10

Query Match 90.3%; Score 28; DB 4; Length 1536;
Best Local Similarity 66.7%; Pred. No. 8.1e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EXPXMR 6
Db 896 FLPAWR 901

RESULT 3
US-09-091-219-25
; Sequence 25, Application US/09091219
; Patent No. 6171592
; GENERAL INFORMATION:
; APPLICANT: STUDDERT, Michael J.
; APPLICANT: CRABB, Brendan S.
; TITLE OF INVENTION: EQUINE RHINOVIRUS 1 PROTEINS
; FILE REFERENCE: 040268/0151
; CURRENT APPLICATION NUMBER: US/09/091,219
; CURRENT FILING DATE: 1998-10-05
; EARLIER APPLICATION NUMBER: PCT/AU96/00815
; EARLIER FILING DATE: 1996-12-18
; EARLIER APPLICATION NUMBER: AU PNT201
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 2232
; TYPE: PRT
; ORGANISM: equine rhinovirus 1
;; US-09-091-219-25

Query Match 90.3%; Score 28; DB 4; Length 2232;
Best Local Similarity 66.7%; Pred. No. 1.1e+03;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EXPXMR 6
Db 2227 FVPTWR 2232

RESULT 4
US-09-091-219-2
; Sequence 2, Application US/09091219
; Patent No. 6171592
; GENERAL INFORMATION:
; APPLICANT: STUDDERT, Michael J.
; APPLICANT: CRABB, Brendan S.
; TITLE OF INVENTION: EQUINE RHINOVIRUS 1 PROTEINS
; FILE REFERENCE: 040268/0151
; CURRENT APPLICATION NUMBER: US/09/091,219
; CURRENT FILING DATE: 1998-10-05
; EARLIER APPLICATION NUMBER: PCT/AU96/00815
; EARLIER FILING DATE: 1996-12-18
; EARLIER APPLICATION NUMBER: AU PNT201
; EARLIER FILING DATE: 1995-12-18
; NUMBER OF SEQ ID NOS: 25

;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 2
;; LENGTH: 2247
;; TYPE: PRT
;; ORGANISM: equine rhinovirus 1
;; US-09-091-219-2

Query Match 90.3%; Score 28; DB 4; Length 2247;
Best Local Similarity 66.7%; Pred. No. 1.2e+03;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EXPXMR 6
Db 2227 FVPTWR 2232

RESULT 5
US-08-215-137-13
; Sequence 13, Application US/08215137
; Patent No. 5614370
; GENERAL INFORMATION:
; APPLICANT: Konteatis, Zenon
; APPLICANT: Siciliano, Salvatore J
; APPLICANT: Springer, Martin S
; TITLE OF INVENTION: ASSAY TO IDENTIFY HUMAN C5A ANTAGONISTS
; TITLE OF INVENTION: AND AGONISTS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Merck & Co., Inc.
; STREET: 126 Lincoln Ave., P.O. Box 2000
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: United States of America
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/215,137
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Bengen, Gerard H
; REGISTRATION NUMBER: 35,746
; REFERENCE/DOCKET NUMBER: 19108
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-3901
; TELEFAX: (908) 594-4720
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: /label= BH
; OTHER INFORMATION: /note= "either the natural phenylalanine amino
; OTHER INFORMATION: terminus or the Bolton-Hunter modified peptide
; OTHER INFORMATION: having the group 3-(p-hydroxyphenyl)propionyl group
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 4
; OTHER INFORMATION: /label= dCha
; OTHER INFORMATION: /note= "D-cyclohexylalanine"

FEATURE: Modified-site
LOCATION: 6
OTHER INFORMATION: /label= darg
OTHER INFORMATION: /note= "D-arginine"
US-08-215-137-13

Query Match 87.1%; Score 27; DB 1; Length 6;
Best Local Similarity 83.3%; Pred. No. 1.6e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FXPXMR 6
1 1 1 1 1
DB 1 FXPXMR 6

RESULT 6
US-09-268-364-21
Sequence 21, Application US/09268364A
Patent No. 6204063
GENERAL INFORMATION:
APPLICANT: Allen, Steve
APPLICANT: Lee, Jian Ming
APPLICANT: Lightner, Jonathan
TITLE OF INVENTION: PLANT GLYCOLYSIS AND RESPIRATION ENZYMES
FILE REFERENCE: BB-1154
CURRENT APPLICATION NUMBER: US/09/268,364A
CURRENT FILING DATE: 1999-03-15
EARLIER APPLICATION NUMBER: 60/079,387
EARLIER FILING DATE: March 16, 1998
NUMBER OF SEQ ID NOS: 21
SOFTWARE: Microsoft Office 97
SEQ ID NO 21
LENGTH: 456
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-268-364-21

Query Match 87.1%; Score 27; DB 4; Length 456;
Best Local Similarity 66.7%; Pred. No. 3.9e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
1 1 1 1 1
DB 209 FXPXMR 214

RESULT 7
US-08-766-858A-27
Sequence 27, Application US/08766858A
Patent No. 5935782
GENERAL INFORMATION:
APPLICANT: Clanciotto, Nicholas P.
APPLICANT: Hickey, Erin K.
APPLICANT: O'Connell, William A.
TITLE OF INVENTION: METHODS AND MATERIALS FOR DETECTING
TITLE OF INVENTION: ECGIONELLA PNEUMOPHILA
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross P.C.
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/766,858A
FILING DATE: 13-DEC-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/011,545
FILING DATE: 13-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Crook, Wanneil M.
REGISTRATION NUMBER: 31,071
REFERENCE/DOCKET NUMBER: 3501-4-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 42 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-766-858A-27

Query Match 83.9%; Score 26; DB 2; Length 42;
Best Local Similarity 50.0%; Pred. No. 64;
Matches 3; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
1 1 1 1 1
DB 20 FXPXMR 25

RESULT 8
US-09-362-831-3
Sequence 3, Application US/09362831
Patent No. 6306400
GENERAL INFORMATION:
APPLICANT: BUBLOT et al.
TITLE OF INVENTION: AVIAN RECOMBINANT LIVE VACCINE USING AS VECTOR, THE
TITLE OF INVENTION: AVIAN INFECTIOUS LARYNGOTRACHEITIS VIRUS
FILE REFERENCE: 454313-2520
CURRENT APPLICATION NUMBER: US/09/362,831
CURRENT FILING DATE: 1999-07-28
NUMBER OF SEQ ID NOS: 28
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 3
LENGTH: 254
TYPE: PRT
ORGANISM: Infectious Laryngotracheitis Virus
US-09-362-831-3

Query Match 83.9%; Score 26; DB 4; Length 254;
Best Local Similarity 66.7%; Pred. No. 3.5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
1 1 1 1 1
DB 186 FXPXMR 191

RESULT 9
US-08-997-080-75
Sequence 75, Application US/08997080
Patent No. 5968524
GENERAL INFORMATION:
APPLICANT: WATSON, JAMES D.
APPLICANT: TAN, PAUL L.J.
TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-
NUMBER OF SEQUENCES: 194
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman

STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/997,080
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000,1007
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 273 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-997-080-75

Query Match 83.9%; Score 26; DB 2; Length 273;
Best Local Similarity 66.7%; Pred. No. 3.8e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPYR 6
DB 267 FDPWMR 272

RESULT 10
US-08-997-362-75
Sequence 75, Application US/08997362
Patent No. 5985287
GENERAL INFORMATION:
APPLICANT: Tan, Paul
APPLICANT: Hiyyama, Jun
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Scott, Linda
APPLICANT: Prestidge, Ross
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
NUMBER OF SEQUENCES: 194
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/997,362
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970
FILING DATE: June 12, 1997
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347
FILING DATE: August 29, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000,1002c2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 273 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-997-362-75

Query Match 83.9%; Score 26; DB 2; Length 273;
Best Local Similarity 66.7%; Pred. No. 3.8e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPYR 6
DB 267 FDPWMR 272

RESULT 11
US-08-873-970-75
Sequence 75, Application US/08873970
Patent No. 6001361
GENERAL INFORMATION:
APPLICANT: Tan, Paul
APPLICANT: Hiyyama, Jun
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Scott, Linda
APPLICANT: Prestidge, Ross
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/873,970
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/705,347
FILING DATE: 29-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000,1002c1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563

TELEX:
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 273 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-873-970-75

Query Match: 83.9%; Score 26; DB 3; Length 273;
Best Local Similarity 66.7%; Pred. No. 3.8e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXMR 6
1 1 1 1
DB 267 FDPWMR 272

RESULT 12

US-09-095-855-75
Sequence 75, Application US/09095855
Patent No. 6160093

GENERAL INFORMATION:

APPLICANT: Tan, Paul
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Prestidge, Ross
TITLE OF INVENTION: Compounds and Methods for
TITL OF INVENTION: Treatment and Diagnosis of Mycobacterial Infections
NUMBER OF SEQUENCES: 208
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/095,855
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/705,347
FILING DATE: 29-AUG-1996
APPLICATION NUMBER: 08/873,970
FILING DATE: 12-JUN-1997
APPLICATION NUMBER: 08/997,362
FILING DATE: 23-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002c3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563

INFORMATION FOR SEQ ID NO: 75:

SEQUENCE CHARACTERISTICS:
LENGTH: 273 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-095-855-75

Query Match 83.9%; Score 26; DB 4; Length 273;
Best Local Similarity 66.7%; Pred. No. 3.8e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXMR 6
1 1 1 1
DB 267 FDPWMR 272

RESULT 13

US-08-997-080-194
Sequence 194, Application US/08997080
Patent No. 5968524

GENERAL INFORMATION:

APPLICANT: TAN, PAUL L.J.
APPLICANT: WATSON, JAMES D.
TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-
NUMBER OF SEQUENCES: 194
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/997,080
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1007
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563

INFORMATION FOR SEQ ID NO: 194:

SEQUENCE CHARACTERISTICS:
LENGTH: 370 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-997-080-194

Query Match 83.9%; Score 26; DB 2; Length 370;
Best Local Similarity 66.7%; Pred. No. 5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXMR 6
1 1 1 1
DB 364 FDPWMR 369

RESULT 14

US-08-997-362-194
Sequence 194, Application US/08997362
Patent No. 5985287

GENERAL INFORMATION:

APPLICANT: Tan, Paul
APPLICANT: Hiyma, Jun
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Scott, Linda

APPLICANT: Prestidge, Ross
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
NUMBER OF SEQUENCES: 194
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/997,362
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970
FILING DATE: June 12, 1997
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347
FILING DATE: August 29, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002c2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ. ID NO: 194:
SEQUENCE CHARACTERISTICS:
LENGTH: 370 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-997-362-194

Query Match 83.9%; Score 26; DB 2; Length 370;
Best Local Similarity 66.7%; Pred. No. 5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPXMR 6
Db 364 FDPWMR 369

RESULT 15
US-09-095-855-194
Sequence 194, Application US/09095855
Patent No. 6160093
GENERAL INFORMATION:
APPLICANT: Tan, Paul
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Prestidge, Ross
TITLE OF INVENTION: Compounds and Methods for
TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
NUMBER OF SEQUENCES: 208
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/095,855
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/705,347
FILING DATE: 28-AUG-1996
APPLICATION NUMBER: 08/873,970
FILING DATE: 12-JUN-1997
APPLICATION NUMBER: 08/997,362
FILING DATE: 23-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002c3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ. ID NO: 194:
SEQUENCE CHARACTERISTICS:
LENGTH: 370 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-095-855-194

Query Match 83.9%; Score 26; DB 4; Length 370;
Best Local Similarity 66.7%; Pred. No. 5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPXMR 6
Db 364 FDPWMR 369

Search completed: February 27, 2002, 11:36:06
Job time: 145 sec

~ Feb 27 12:13:46 2002

us-09-446-109a-12.ra1

Page 7

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:41:17 (Search time 303.5 Seconds
(without alignments)
1.220 Million cell updates/sec)

Title: US-09-446-109a-13
Perfect score: 30
Sequence: 1 FPXMR 5

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

- 1: /SIDSR/gcgdata/geneseq/AA1980.DAT.*
- 2: /SIDSR/gcgdata/geneseq/AA1981.DAT.*
- 3: /SIDSR/gcgdata/geneseq/AA1982.DAT.*
- 4: /SIDSR/gcgdata/geneseq/AA1983.DAT.*
- 5: /SIDSR/gcgdata/geneseq/AA1984.DAT.*
- 6: /SIDSR/gcgdata/geneseq/AA1985.DAT.*
- 7: /SIDSR/gcgdata/geneseq/AA1986.DAT.*
- 8: /SIDSR/gcgdata/geneseq/AA1987.DAT.*
- 9: /SIDSR/gcgdata/geneseq/AA1988.DAT.*
- 10: /SIDSR/gcgdata/geneseq/AA1989.DAT.*
- 11: /SIDSR/gcgdata/geneseq/AA1990.DAT.*
- 12: /SIDSR/gcgdata/geneseq/AA1991.DAT.*
- 13: /SIDSR/gcgdata/geneseq/AA1992.DAT.*
- 14: /SIDSR/gcgdata/geneseq/AA1993.DAT.*
- 15: /SIDSR/gcgdata/geneseq/AA1994.DAT.*
- 16: /SIDSR/gcgdata/geneseq/AA1995.DAT.*
- 17: /SIDSR/gcgdata/geneseq/AA1996.DAT.*
- 18: /SIDSR/gcgdata/geneseq/AA1997.DAT.*
- 19: /SIDSR/gcgdata/geneseq/AA1998.DAT.*
- 20: /SIDSR/gcgdata/geneseq/AA1999.DAT.*
- 21: /SIDSR/gcgdata/geneseq/AA2000.DAT.*
- 22: /SIDSR/gcgdata/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	96.7	10	AAV08336	O. hannah venom al
2	29	96.7	11	AAV08334	O. hannah venom al
3	29	96.7	38	AAV58491	Blinding domain of
4	29	96.7	94	AAV06599	Rat sodium channel
5	29	96.7	94	AAV20130	Rat truncated sodi
6	29	96.7	135	AAV04918	Mycobacterium spec
7	29	96.7	177	AAV58892	Breast and ovarian
8	29	96.7	299	AAV52488	Helicobacter pylori
9	29	96.7	314	AAV93943	CD46 construct del
10	29	96.7	324	AAV15233	CD46 from clone pm
11	29	96.7	359	AAV70396	Cellulase. Cellul

Result No.	Score	Query Match	Length	ID	Description
12	29	96.7	372	AAV27793	New platelet facto
13	29	96.7	372	AAV92239	Chemokine superfam
14	29	96.7	372	AAV68813	Human lymphocyte P
15	29	96.7	372	AAV06644	Human Burkitt's Ly
16	29	96.7	372	AAV90627	Human G protein-co
17	29	96.7	372	AAV90661	Human mutant G pro
18	29	96.7	373	AAV15230	CD46 from clone pm
19	29	96.7	376	AAV93939	CD46 wild-type. H
20	29	96.7	377	AAV15229	CD46 from clone pm
21	29	96.7	377	AAV93942	CD46 construct sub
22	29	96.7	377	AAV93941	CD46 construct sub
23	29	96.7	377	AAV27484	Human MCP. Homo s
24	29	96.7	384	AAV10924	Human membrane cof
25	29	96.7	384	AAV86316	Human CD46. Homo
26	29	96.7	418	AAV13589	Streptomyces globi
27	29	96.7	421	AAV58394	Lung cancer associ
28	29	96.7	421	AAV75528	Human colon cancer
29	29	96.7	497	AAV04972	Mycobacterium spec
30	29	96.7	520	AAV90739	C glutamicum prote
31	29	96.7	589	AAV34524	Porphyromonas ging
32	29	96.7	590	AAV94928	Human secreted pro
33	29	96.7	593	AAV34523	Porphyromonas ging
34	29	96.7	606	AAV34393	Porphyromonas ging
35	29	96.7	611	AAE03762	CAB-2 chimeric pro
36	29	96.7	612	AAV94498	Human protein sequ
37	29	96.7	614	AAV20991	H. pylori inner me
38	29	96.7	621	AAV94049	Human protein sequ
39	29	96.7	634	AAV40125	Human polypeptide
40	29	96.7	634	AAV92742	Human protein sequ
41	29	96.7	634	AAV93478	Human protein sequ
42	29	96.7	634	AAV95120	Human protein sequ
43	29	96.7	1464	AAV79294	An antigen from de
44	28	93.3	42	AAV15108	Peptide #1542 enco
45	28	93.3	42	AAV27563	Peptide #1600 enco

ALIGNMENTS

RESULT 1

AAV08336 standard; Peptide: 10 AA.

AC AAV08336:

DT 19-JUL-1999 (first entry)

XX

XX O. hannah venom alpha-neurotoxin homologous peptide 5.

XX Snake; venom; King cobra; alpha-neurotoxin; toxin; analgesia; screening.

XX Ophiophagus hannah.

XX W09924055-A1.

XX 20-MAY-1999.

XX 03-NOV-1998: 98MO-SG00087.

XX 06-NOV-1997: 97SG-0003972.

XX (UYST-) UNIV SINGAPORE NAT.

XX copaladishnakone P, Gwee MCE, Klni RM, Pu XC, Wong PT:

XX WPI: 1999-327205/27.

XX Snake neurotoxin derived peptides

XX Claim 11: Page 43: 46pp: English.

XX This invention describes novel peptide derivatives of alpha-neurotoxin, derived from snake venom toxins of Ophiophagus hannah and which are

CC capable of inducing or facilitating analgesia. The peptides, optionally
 CC with a carrier and/or diluent, are used to induce analgesia in a subject.
 CC Antibodies derived from the products of the invention can be used to
 CC screen for subject peptides, e.g. for purification purposes.

XX
 SQ Sequence 10 AA:

Query Match 96.7%: Score 29; DB 20; Length 10;
 Best Local Similarity 80.0%: Pred. No. 13;
 Matches 4: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 11 11
 Db 3 fplwr 7

RESULT 2

AAV08334
 ID AAV08334 standard; Peptide: 11 AA.

XX
 AC AAV08334:

XX
 DT 19-JUL-1999 (first entry)

XX
 DE O. hannah venom alpha-neurotoxin homologous peptide 3.

XX
 KW Snake: venom: King cobra; alpha-neurotoxin; toxin; analgesia; screening.

XX
 OS Ophiophagus hannah.

XX
 PN MO9924055-A1.

XX
 PD 20-MAY-1999.

XX
 PE 03-NOV-1998; 98WO-SC00087.

XX
 PR 06-NOV-1997; 97SG-0003972.

XX
 PA (UYST-) UNIV SINGAPORE NAT.

XX
 PI Copaladishnakone P, Gwee MCE, Kini RM, Pu XC, Wong PT;

XX
 DR WPI: 1999-327205/27.

XX
 PT Snake neurotoxin derived peptides

XX
 PS Claim 9: Page 42; 46pp; English.

XX
 CC This invention describes novel peptide derivatives of alpha-neurotoxin,
 CC derived from snake venom toxins of Ophiophagus hannah and which are
 CC capable of inducing or facilitating analgesia. The peptides, optionally
 CC with a carrier and/or diluent, are used to induce analgesia in a subject.
 CC Antibodies derived from the products of the invention can be used to
 CC screen for subject peptides, e.g. for purification purposes.

XX
 SQ Sequence 11 AA:

Query Match 96.7%: Score 29; DB 20; Length 11;
 Best Local Similarity 80.0%: Pred. No. 14;
 Matches 4: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 11 11
 Db 3 fplwr 7

RESULT 3

AAV58491
 ID AAV58491 standard; Protein: 38 AA.

XX
 AC AAV58491:

XX
 DT 18-APR-1995 (first entry)

XX
 DE Binding domain of a polystyrene-binding TSAR (SB.9-5).

XX
 KW TSAR: totally synthetic affinity reagent; synthetic; binding domain;
 KW effector domain; concatenated heterofunctional protein; linker;
 KW direct; rapid; detection; screening; treatment; monoclonal antibody;
 KW Sm antigen; mouse model; autoimmune disease; SLE; MAb;
 KW systemic lupus erythematosus; polystyrene; SB.

XX
 OS Synthetic.

XX
 PN MO9418318-A.

XX
 PD 18-AUG-1994.

XX
 PE 01-FEB-1994; 94WO-US00977.

XX
 PR 01-FEB-1993; 93US-0013416.

XX
 PR 30-DEC-1993; 93US-0176500.

XX
 PR 31-JAN-1994; 94US-0189331.

XX
 PA (UYNC-) UNIV NORTH CAROLINA.

XX
 PT Fowlkes DM, Kay BK;

XX
 DR WPI: 1994-279739/34.

XX
 IT Identifying proteins or peptide(s) which bind a ligand - by
 IT screening a recombinant vector library expressing fusion proteins
 IT comprising a binding domain and an effector domain

XX
 PS Example 7.8: Page 120; 255pp; English.

XX
 CC AAV58491-507 are amino acid sequences of the binding domain of TSAR
 CC (Totally Synthetic Affinity Reagents) peptides. These particular
 CC peptides are encoded by the polystyrene (SB) binding phage. There
 CC is no apparent linear motif, but the peptides are rich in Trp, Tyr and
 CC Gly, poor in Arg, Val and Lys, and completely lack cysteine residues.
 CC The non-variable amino acids at the NH2 and COOH terminals are
 CC not shown. TSAR peptides are generated using generic oligonucleotides
 CC (see AAV70470-73 for examples). TSARs are concatenated heterofunctional
 CC proteins or peptides, comprising at least two functional regions - a
 CC binding domain with affinity for a ligand and a second effector peptide
 CC portion that is chemically or biologically active. They may further
 CC comprise a linker peptide between the 2 domains. The TSARs or compms.
 CC comprising a TSAR binding domain can be used in vivo to deliver a
 CC chemically or biologically active moiety, eg. metal ion, radioisotope,
 CC peptide, toxin or enzyme, to the specific target or on the cell. They
 CC can also replace the function of macromolecules eg. monoclonal or
 CC polyclonal antibodies and therefore circumvent the need for complex
 CC methods of hybridoma formation or in vivo antibody production. The TSARs
 CC are easily characterised and have designed activity allowing direct and
 CC rapid detection in a screening process.

XX
 SQ Sequence 38 AA:

Query Match 96.7%: Score 29; DB 15; Length 38;
 Best Local Similarity 80.0%: Pred. No. 46;
 Matches 4: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 11 11
 Db 10 fplwr 14

RESULT 4

AAV06599
 ID AAV06599 standard; Protein: 94 AA.

XX
 AC AAV06599:

```

XX 26-OCT-1999 (first entry)
XX
XX Rat sodium channel Nan splice variant C-terminal sequence.
DE
XX NaN; sodium channel; ion transport; rat; dorsal root ganglia;
KM pain; paraesthesia; hyperexcitability; therapy.
XX
XX Rattus sp.
XX
XX MO9938889-A2.
XX
XX 05-AUG-1999:
XX
XX 29-JAN-1999: 99MO-US02008.
XX
XX 20-NOV-1998: 98US-0109402.
XX
XX 29-JAN-1998: 98US-0072990.
XX
XX (UYVA ) UNIV YALE.
XX
XX Dlb-Hajj S, Waxman S;
XX
XX WPI: 1999-479168/40.
XX
XX New isolated nucleic acids encoding sodium channels, used to develop
PT products for treating acute or chronic pain or hyperexcitability
PT phenomena
XX
XX Example 5; page 35; 91pp; English.
XX
XX This is the C-terminal sequence of a splice variant of rat sodium
CC channel NaN. This truncated variant comprises amino acids 1-1378
CC of the Nan sequence given in AAY06596 and the present, 94-amino acid
CC sequence at the C-terminal end. It thus lacks the 387 C-terminal
CC amino acids of full-length NaN. The variant arises from the use
CC of a cryptic donor splice site in exon 23 and a novel exon 23'
CC which is located in intron 23 of the rat NaN gene. NaN is a
CC previously unidentified voltage-gated sodium channel protein that
CC is preferentially expressed in dorsal root ganglia or trigeminal
CC ganglia, and which produces a TRX-R sodium current. The invention
CC provides methods for identifying agents that modulate NaN channel
CC activity or expression and for using such agents to treat acute or
CC chronic pain, paraesthesia and hyperexcitability phenomena.
XX
XX Sequence 94 AA:
SQ

```

```

Query Match 96.7%: Score 29; DB 20; Length 94;
Best Local Similarity 80.0%: Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Oy 1 PPXWR 5
   1 1 1
Db 52 fpawr 56

```

```

RESULT 5
AAB20130
ID AAB20130 standard; Protein: 94 AA.
XX
XX AAB20130:
XX
XX 30-APR-2001 (first entry)
XX
XX Rat truncated sodium channel C-terminal region.
XX
XX Sodium channel; NaN; rat; tetrodotoxin resistant; pain;
KM paraesthesia; hyperexcitability; analgesic; vaccine; therapy;
XX diagnosis.
XX
XX Rattus norvegicus.
XX

```

```

PN MO200105831-A1.
XX
XX 25-JAN-2001.
XX
XX 14-JUL-2000: 2000MO-US19342.
XX
XX 16-JUL-1999: 99US-0354147.
XX
XX (UYVA ) UNIV YALE.
XX
XX Dlb-Hajj S, Waxman SG;
XX
XX WPI: 2001-103147/11.
XX
XX Nucleic acid molecules encoding human tetrodotoxin resistant sodium
PT channels, useful for preventing, diagnosing and treating pain.
PT paraesthesia and/or hyperexcitability phenomena -
XX
XX Example 5; page 34; 162pp; English.
XX
XX The present sequence is that of the C-terminal portion of a
CC truncated rat NaN variant. The variant lacks the 387 C-terminal
CC amino acids of full-length rat NaN (see AAB20122), which are
CC replaced by the present 94-amino acid polypeptide. The N-terminal
CC 1378 amino acids of the truncated variant are identical to those
CC of full-length NaN. The new sequence arises from the use of a
CC cryptic donor splice site in exon 23 and a novel exon 23' which is
CC located in intron 23. Rat NaN is a novel tetrodotoxin resistant
CC sodium channel NaN belonging to the a-subunit voltage-gated sodium
CC channel protein family. These sodium channels underlie the
CC generation and propagation of impulses in excitable cells such as
CC neurons and muscle fibres. Preferential expression of NaN on
CC sensory neurons innervating the body (dorsal root ganglia) and
CC the face (trigeminal ganglia), but not on other neurons, makes
CC it a very useful target for diagnostic and/or therapeutic uses in
CC relation to acute and/or chronic pain pathologies. A claimed
CC method of treating pain, paraesthesia and/or hyperexcitability
CC phenomena in a human or animal subject involves administering an
CC agent that alters sodium current flow through NaN channels, or
CC which modulates transcription or translation of NaN mRNA, in
CC dorsal root ganglia or trigeminal neurons. NaN polypeptides can
CC be obtained by recombinant expression, and used to treat disorders
CC associated with decreased sodium channel expression, to screen for
CC compounds that modulate sodium channel expression or activity,
CC and to raise antibodies useful as diagnostic agents.
XX
XX Sequence 94 AA:
SQ

```

```

Query Match 96.7%: Score 29; DB 22; Length 94;
Best Local Similarity 80.0%: Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Oy 1 PPXWR 5
   1 1 1
Db 52 fpawr 56

```

```

RESULT 6
AAY04918
ID AAY04918 standard; Protein: 135 AA.
XX
XX AAY04918:
XX
XX 06-JUL-1999 (first entry)
XX
XX Mycobacterium species protein sequence 358.
XX
XX Secreted protein: Mycobacterium; primer; PCR; amplification; probe;
KM hybridisation; detection; vaccine; immunisation; infection.
XX
XX Mycobacterium sp.
XX

```

PN MO9090186-A2.
 XX
 PD 25-FEB-1999.
 XX
 PE 14-AUG-1998: 98MO-FR01813.
 XX
 PR 11-SEP-1997: 97FR-0011325.
 PR 14-AUG-1997: 97FR-0010404.
 XX
 PA (INSP) INST PASTEUR.
 XX
 PI Cicaquel B, Llm EM, Pellicic V, Portnoi D, Goguett de la Salmoniere Y;
 PI Cigueno A;
 XX
 DR WPI: 1999-181045/15.
 DR N-PSDB: AAX341171.
 XX
 PT Mycobacterial DNA vectors containing reporter constructs - for
 PT identifying coding or promoter sequences involved in
 PT infection-associated protein expression
 PS
 PS Claim 32: Fig 35B: 309pp: French.
 XX
 CC Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted
 CC proteins from various Mycobacterium species microorganisms. The
 CC encoding nucleotide sequences can be used as primers and probes for
 CC methods for detecting and identifying mycobacteria, especially belonging
 CC to the M. tuberculosis complex. The encoded proteins can be used in
 CC vaccines for immunisation against a bacterial or viral infection.
 CC
 SQ Sequence 135 AA:

Query Match 96.7%; Score 29; DB 20; Length 135;
 Best Local Similarity 80.0%; Pred. No. 1.5e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 Db 66 fpawr 70

RESULT 7
 AAB58892
 ID AAB58892 standard: Protein: 177 AA.
 AC AAB58892:
 XX
 DT 27-MAR-2001 (first entry)
 XX
 DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 600.
 XX
 KW Human: breast cancer; ovarian cancer; cytostatic; immunosuppressive;
 KW nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 KW antidiabetic; antiinflammatory; anticancer; vulnery; anticonvulsant;
 KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
 KW Addison's disease; allergy; autoimmune haemolytic anaemia;
 KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
 KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
 KW cardiovascular disorder; wound healing; neurological disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200055173-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000: 2000MO-US05881.
 XX
 PR 12-MAR-1999: 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX

PI Rosen CA, Ruben SM;
 XX
 DR WPI: 2000-611515/58.
 DR N-PSDB: AAF21795.
 XX
 PT New human breast and ovarian cancer associated gene sequences and the
 PT polypeptides encoded by these genes, useful in the prevention,
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular
 PT disorders and neurological diseases -
 XX
 PS Claim 11: Page 1037-1038; 1299pp: English.
 XX
 CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
 CC associated with breast and ovarian cancer. Included in the invention are
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
 CC isolation and characterisation of the DNA and protein sequences of the
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist
 CC or antagonist sequences exhibit cytostatic; immunosuppressive;
 CC nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 CC antidiabetic; antiinflammatory; anticancer; vulnery; anticonvulsant;
 CC antibacterial; antifungal; antiparasitic and cardiant activity. The
 CC polynucleotide and protein sequences are used in the diagnosis of cancer,
 CC particularly breast and ovarian cancer. The nucleic acid sequences,
 CC proteins, agonists and antagonists may also be used in the diagnosis,
 CC prevention and treatment of immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; cardiovascular disorders such as
 CC myocardial ischaemias; wound healing; neurological diseases such as
 CC cerebral anoxia and epilepsy; and infectious diseases.
 CC
 SQ Sequence 177 AA:

Query Match 96.7%; Score 29; DB 21; Length 177;
 Best Local Similarity 80.0%; Pred. No. 2e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 Db 22 fpawr 26

RESULT 8
 AAB52488
 ID AAB52488 standard: Protein: 299 AA.
 AC AAB52488:
 XX
 DT 23-FEB-2001 (first entry)
 XX
 DE Helicobacter pylori bait polypeptide #6.
 XX
 KW Helicobacter pylori; two-hybrid system; protein-protein interaction;
 KW bait polypeptide; gastric ulcer; antibacterial.
 XX
 OS Helicobacter pylori.
 XX
 PN WO200066722-A1.
 XX
 PD 09-NOV-2000.
 XX
 PF 14-APR-2000: 2000MO-IB00603.
 XX
 PR 30-APR-1999: 99EP-0401066.
 XX
 PA (HYBR-) HYBRIGENICS SA.
 XX
 PT Legrain P, Selig L, Rain J;
 XX
 DR WPI: 2000-687535/67.
 DR N-PSDB: AAC97234.

XX A two-hybrid system for identifying compounds useful in the treatment
PT of e.g. gastric ulcers comprises producing a collection of recombinant
PT cell clones -
XX
PS Example 5; Page 88-89; 267pp; English.
XX
CC The present sequence is a bait polypeptide used in a Helicobacter
CC pylori two-hybrid screen to identify recombinant cell clone expressing a
CC The method is used to identify a recombinant cell clone expressing a
CC prey polypeptide which is capable of interacting with the bait
CC polypeptide. The two hybrid system is useful for screening compounds
CC for antibacterial activity. It may be used in the treatment of gastric
CC ulcers. The polynucleotides are useful as amplification primers or
CC specific detection probes. The polypeptides, vectors or host cells can
CC be used as immunogens to produce mono- or polyclonal antibodies. The
CC polynucleotides, polypeptides, antibodies, vectors, host cells or
CC modulating agents can be used to produce a pharmaceutical composition.
XX
SQ Sequence 299 AA:

Query Match 96.7%; Score 29; DB 21; Length 299;
Best Local Similarity 80.0%; Pred. No. 3.3e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 1 FPXWR 5
Db 89 fpswr 93

RESULT 9
ID AAR93943 standard; Protein: 314 AA.
XX
XX AAR93943:
XX
XX 21-MAY-1996 (first entry)
XX
XX
XX CD46 construct delSCR2/subSCR3.
XX
XX CD46: recombinant protein; short consensus repeat; SCR:
XX regulator of complement activation; transgenic animal; pig;
XX organ transplantation.
XX
XX Synthetic.
XX
XX WO9606937-A1.
XX
XX 07-MAR-1996.
XX
XX 30-AUG-1995: 95WO-AU00553.
XX
XX 30-AUG-1994: 94AU-0007724.
XX
XX (AUST-) AUSTIN RES INST.
XX
XX Christiansen D, Loveland B, McKenzie IFC, Milland J;
XX
XX MPI: 1996-160368/16.
XX
XX N-PSDB: AAT17599.
XX
XX Increasing prodn. of recombinant proteins, esp. CD46 - by reducing
XX the amt. of A and/or T in an A and/or T rich region of encoding gene
XX exon
XX
XX Claim 12; Page 40-41; 60pp; English.
XX
XX CD46 delSCR2/subSCR3 (AAR93943) is the product of a cDNA construct
XX (AAT17599) obtd. by splice overlap extension PCR of wild-type CD46
XX cDNA (AAT17595). The A+T content of A+T-rich exon 5 of the gene,
XX encoding the short consensus repeat (SCR) 3 of CD46, was lowered
XX and SCR2 was deleted. This resulted in improved prodn. in

CC eukaryotic host cells, e.g. CHO-K1 and COS-7. CD46 delSCR2/subSCR3
CC is used to prevent complement- or inflammation-mediated tissue
CC damage, to improve immunity to tumours or viruses, to control
CC fertilisation and to prevent spontaneous abortion. Expression in
CC transgenic animals, esp. pigs, provides organs suitable for
CC transplantation.
XX
SQ Sequence 314 AA:

Query Match 96.7%; Score 29; DB 17; Length 314;
Best Local Similarity 80.0%; Pred. No. 3.4e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 1 FPXWR 5
Nb 11 fpswr 15

RESULT 10
ID AAR15233 standard; Protein: 324 AA.
XX
XX AAR15233:
XX
XX 17-MAR-1992 (first entry)
XX
XX
XX CD46 from clone pm5.8.
XX
XX SCR: short consensus repeats; transmembrane; cytoplasmic;
XX membrane cotactor protein; MCP.
XX
XX Homo sapiens.
XX
XX
XX Key Location/Qualifiers
XX Peptide 1..34
XX Protein /label= sig-peptide
XX /label= 34..324
XX /label= mat_protein
XX Modified-site 83
XX /label= N-glycosylation_site
XX Modified-site 114
XX /label= N-glycosylation_site
XX Domain 289..304
XX /label= hydrophobic_transmembrane_domain
XX
XX WO9118097-A.
XX
XX 28-NOV-1991.
XX
XX 10-MAY-1991: 91WO-AU00199.
XX
XX 11-MAY-1990: 90AU-0000133.
XX
XX (UYME-) UNIV MELBOURNE.
XX
XX Purcell DET, Russell SM, McKenzie JFC;
XX
XX MPI: 1991-369251/50.
XX
XX N-PSDB: AAQ14919.
XX
XX New CD46 membrane co:factor protein variants - useful as probes
XX to identify CD46 isoforms and for diagnosing spontaneous
XX abortion, inhibiting immuno:response and treating Leukemia
XX
XX Disclosure; Fig 1A and 3B; 77pp; English.
XX
XX The sequence of the pm5.8 clone is identical to the pm5.1 clone in
XX the portion encoding the NH2 leader and four SCR regions. The
XX sequence after nucleotide 890 was different, and results from
XX reading through of the cDNA into an intron sequence after the
XX fourth SCR. The protein encoded by this new sequence encodes
XX a 16 amino acid hydrophobic region, probably serving as a membrane

CC spanning sequence. The putative cytoplasmic tail would also
 CC contain a high proportion of charged amino acids.
 CC See also AAQ14915-25, AAQ15211-12 and AAR15457-59.
 XX
 SQ Sequence 324 AA;

Query Match 96.7%: Score 29; DB 12; Length 324;
 Best Local Similarity 80.0%: Pred. No. 3.5e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
 11 11
 Db 11 fgsr 15

RESULT 11

AAP70396
 ID AAP70396 standard; Protein: 359 AA.

XX
 AC AAP70396;

XX
 DT 30-APR-1991 (first entry)

XX
 DE Cellulase.

XX
 KM Cellulase: plasmid; transformation.

XX
 OS Cellulomonas uda CB4 (FER.

XX
 PN JP62175178-A.

XX
 PD 31-JUL-1987.

XX
 PF 30-JAN-1986; 86JP-0016800.

XX
 PR 30-JAN-1986; 86JP-0016800.

XX
 PA (ENER-) SHIN ENERGY SOGO KAIHATSU KIKO.

XX
 (NENR-) NENRYOY ALCOHOL KAIHATSU CIJUTSU KENKYU KUMIAT.

XX
 DR WPI: 1987-253827/36.

XX
 DR N-PSDB: AAN70617.

XX
 PT Bio-engineering cellulase prodn. - by preparing DNA chain,

XX
 PT preparing plasmid contg. DNA chain in the gene expressible state,

XX
 PT and transforming bacterial by plasmid

XX
 PS Disclosure; Fig 1(A-B): 12pp; Japanese.

XX
 CC Transformation E.coli JM103 (FERM P-8612) with the gene encoding

XX
 CC this protein allows expression of the protein and easy and cheap

XX
 CC prodn. of cellulase.

XX
 SQ Sequence 359 AA;

OY 1 FPXWR 5
 11 11
 Db 178 fgsr 182

RESULT 12

AAR27793
 ID AAR27793 standard; Protein: 372 AA.

XX
 AC AAR27793;

XX
 DT 12-MAR-1993 (first entry)

XX
 DE New platelet factor 4 receptor superfamily member PF4AR1.
 XX
 XX
 KM IL-8R; G-protein coupled receptor family; rhodopsin superfamily;
 KM pro-inflammatory cytokine; 8tr.9.
 XX
 OS Homo sapiens.

XX
 PN W09217497-A.

XX
 PD 15-OCT-1992.

XX
 PF 23-MAR-1992; 92MO-US02317.

XX
 PR 29-MAR-1991; 91US-0677211.

XX
 PR 19-DEC-1991; 91US-0810782.

XX
 PA (GETH) GENENTECH INC.

XX
 PI Holmes WE, Lee J, Wood WI;

XX
 DR WPI: 1992-366191/44.

XX
 DR N-PSDB: AAQ37107.

XX
 PT Isolated human platelet factor 4 super-family receptor

XX
 PT polypeptide and corresp. antibodies and DNA - useful as

XX
 PT diagnostic and screening agents, and for treating inflammation or

XX
 PT PF4AR-mediated disorders

XX
 PS Claim 7; Fig 5; 78pp; English.

XX
 CC The IL-8 receptor cDNA sequence was isolated (see AAQ29505) and a

XX
 CC 874bp sub-fragment of the coding sequence was used as a probe to

XX
 CC screen human cell line HL60 and human peripheral blood lymphocyte

XX
 CC cDNA libraries. Two new gene sequences were found that are clearly

XX
 CC related to the IL-8 receptor. One of these was contained in clone

XX
 CC 8tr.9 and is predicted to encode an amino acid sequence which is

XX
 CC 36% and 38% identical with the high and low affinity IL-8 receptor

XX
 CC sequences, respectively. See also AAQ37107.

XX
 SQ Sequence 372 AA;

OY 1 FPXWR 5
 11 11
 Db 352 fgsr 355

XX
 PD 26-MAR-1996 (first entry)

XX
 DE Chemokine superfamily receptor.

XX
 KM Interleukin; IL-8; inflammation; psoriasis; dermatitis;

XX
 KM rheumatoid arthritis; inflammatory bowel disease;

XX
 KM chronic lung inflammation; treatment; antibody;

XX
 KM affinity purification; detection.

XX
 OS Homo sapiens.

XX
 PN US5440021-A.

XX
 PD 08-AUG-1995.

PF 29-MAR-1991: 91US-0677211.
 XX
 PR 25-FEB-1994: 94US-0202056.
 PR 29-MAR-1991: 91US-0677211.
 XX
 PA (CHUN/) CHUNTHARAPAL A.
 PA (HEBE/) HEBERT C.
 PA (KIMK/) KIM K J.
 PA (LEEJ/) LEE J.
 XX
 PI Chuntharapal A, Hebert C, Kim KJ, Lee J:
 XX
 DR WPI: 1995-283151/37.
 DR N-PSDB: AA099009.
 XX
 PT New antibodies against Interleukin 8 type B receptor - used to treat
 PT or prevent inflammation, also for detecting receptor expression and
 PT purification.
 XX
 PS Example 2: Columns 49-52: 62pp; English.
 XX
 CC Antibodies directed against the Interleukin-8 receptor B can be used
 CC to treat or prevent inflammation e.g. psoriasis, dermatitis,
 CC rheumatoid arthritis and particularly inflammatory bowel disease and
 CC chronic lung inflammation. When immobilised, these antibodies may
 CC be used to detect Interleukin-8 receptor B expression in cells and
 CC tissues and for affinity purification of Interleukin-8 receptor B
 CC from cells. This sequence is an additional chemokine superfamily
 CC receptor which was identified by probing lambda libraries of genomic
 CC DNA from a human monocyte-like cell line (U-60) and human peripheral
 CC blood lymphocytes using a large fragment of the Interleukin-8 type
 CC A receptor DNA (See AA099006).
 CC
 XX Sequence 372 AA:
 SQ

Query Match 96.7%: Score 29: DB 16: Length 372:

Best Local Similarity 80.0%: Pred. No. 4e+02: 1: Indels 0: Gaps 0:
Matches 4: Conservative 0: Mismatches 1:

Oy 1 FPXWR 5
 11 11
 Db 352 fpswr 356

RESULT 14

AA068813 ID AAR68813 standard: Protein: 372 AA.

AC AAR68813:

DT 18-JUL-1995 (first entry)

DE Human lymphocyte PF4AR.

XX Interleukin-8 receptor; IL-8 receptor; PF4AR;

KM platelet factor superfamily receptor; lymphocyte; chemotactic;
 KM inflammation; inflammatory disease; arthritis; emphysema; cystic;
 KM fibrosis; colitis; bronchitis; meningitis; therapeutic.

XX Homo sapiens.

OS WO9428931-A.

PN 22-DEC-1994.

XX 07-JUN-1994: 94WO-US06380.

XX 11-JUN-1993: 93US-0076093.

XX (CETI) GENENTECH INC.

PA Chuntharapal A, Hebert C, Kim KJ, Lee J:
 XX
 PI

XX
 DR WPI: 1995-036114/05.
 DR N-PSDB: AA080522.
 XX
 PT Treatment of inflammatory disorders - by administering an
 PT antibody capable of binding a platelet factor 4 superfamily
 PT receptor polypeptide
 XX
 PS Disclosure: Page 56-58: 83pp; English.
 XX
 CC 2 PF4AR members were identified by probing lambda libraries from
 CC human monocyte-like cell line HL-60 and human peripheral blood
 CC lymphocytes using a large fragment of IL-8 receptor DNA (full
 CC sequence given in AA080520). The nucleotide sequences of the 2
 CC PF4ARs are given in AA080521 and AA080522, and their respective
 CC amino acid sequences in AAR68812 and AAR68813.
 CC
 XX Sequence 372 AA:
 SQ

Query Match 96.7%: Score 29: DB 16: Length 372:

Best Local Similarity 80.0%: Pred. No. 4e+02: 1: Indels 0: Gaps 0:
Matches 4: Conservative 0: Mismatches 1:

Oy 1 FPXWR 5
 11 11
 Db 352 fpswr 356

RESULT 15

AA066644 ID AAY06644 standard: Protein: 372 AA.

AC AAY06644:

DT 26-OCT-1999 (first entry)

DE Human Burkitt's lymphoma receptor 1 (BLR1).

XX Burkitt's lymphoma receptor 1; BLR1; human;

KM B lymphocyte chemottractant; BLC; chemokine; ligand;
 KM drug screening; leukaemia; autoimmune disease; therapy.

XX Homo sapiens.

OS WO9928468-A1.

PN 10-JUN-1999.

XX 02-DEC-1998: 98WO-US25561.

XX 02-DEC-1997: 97US-0982493.

XX (REGC) UNIV CALIFORNIA.

XX Cyster JC, Gunn MD, Williams LT;

XX WPI: 1999-493764/A1.

DR N-PSDB: AAX67710.

XX Modulating interaction of a Burkitt's Lymphoma Receptor 1
 PT polypeptide and ligand, useful in drug screens

PS Claim 1: Page 38: 42pp; English.

CC This sequence represents human Burkitt's lymphoma receptor 1
 CC (BLR1). The invention relates to methods for modulating the
 CC interaction of BLR1 with its ligand, B lymphocyte chemottractant
 CC (BLC, see AAY06642). The methods comprise combining BLR1 and BLC
 CC polypeptides with a candidate modulator agent under conditions
 CC whereby, but for the presence of the agent, the polypeptides engage
 CC in a first interaction, and determining a second interaction of the
 CC polypeptides in the presence of the agent, wherein a difference

CC between the first and second interactions indicates that the agent
CC modulates the interaction of the polypeptides. The modulator is
CC preferably an antagonist, especially dominant negative, form of
CC BIC. BIC and BIC agonists and antagonists may be useful for
CC treating viral (e.g. HIV) infections, lymphoma, B lineage
CC leukemia, and autoimmune diseases such as rheumatoid arthritis,
CC thyroiditis and diabetes.

XX
SQ Sequence 372 AA;

Query Match: 96.7%; Score 29; DB 20; Length 372;

Best Local Similarity 80.0%; Pred. No. 4e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXWR 5

Db 352 fpswr 356

Search completed: February 27, 2002, 11:41:18
Job time: 457 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:45:21 : Search time 145.23 Seconds
(without alignments)
2.623 Million cell updates/sec

Title: US-09-446-109a-1.3

Perfect score: -30

Sequence: 1 FPXMR 5

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	96.7	73	1 N20H1	long neurotoxin 1
2	29	96.7	113	2 T07855	translational elonga
3	29	96.7	157	2 T24380	hypothetical prote
4	29	96.7	245	2 D82680	conserved hypothet
5	29	96.7	256	2 D85510	probable EC 3.5. a
6	29	96.7	259	2 A29831	heart-labile entero
7	29	96.7	273	2 T05454	hypothetical prote
8	29	96.7	292	2 C82960	hypothetical prote
9	29	96.7	327	2 S56162	MDCR15 protein - h
10	29	96.7	332	2 JT0585	minor endogluccas
11	29	96.7	349	2 C02913	sperm CD45 - human
12	29	96.7	359	2 I40696	endogluccase - Ce
13	29	96.7	369	2 I57998	membrane cofactor
14	29	96.7	372	2 S26677	G protein-coupled
15	29	96.7	377	2 I54479	membrane cofactor
16	29	96.7	384	2 S01896	membrane cofactor
17	29	96.7	417	2 D86251	protein F25C20.9 l
18	29	96.7	479	2 T35441	aldohyde dehydroge
19	29	96.7	497	2 D83628	probable aldehyde
20	29	96.7	555	2 F69312	heterodisulfide re
21	29	96.7	625	2 H70330	hypothetical prote
22	29	96.7	642	2 D71909	ferrous iron trans
23	29	96.7	642	2 G64605	iron(II) transport
24	29	96.7	650	2 T22002	hypothetical prote
25	29	96.7	735	2 A83006	hypothetical prote
26	29	96.7	1451	2 I40325	dermatonecrotic tox
27	28	93.3	68	2 S75058	transposase ssr289
28	28	93.3	84	2 S75091	hypothetical prote
29	28	93.3	84	2 S76443	hypothetical prote

30	28	93.3	93	2 S75008	transposase ss1192
31	28	93.3	103	2 T47718	hypothetical prote
32	28	93.3	114	2 S77061	transposase sl1066
33	28	93.3	143	1 HKM48	heat shock protein
34	28	93.3	143	1 HKM41	heat shock protein
35	28	93.3	147	2 A71560	hypothetical prote
36	28	93.3	149	2 T35846	probable integrat
37	28	93.3	157	2 F81710	conserved hypothet
38	28	93.3	169	2 C69300	conserved hypothet
39	28	93.3	254	2 S76459	transposase sl1186
40	28	93.3	254	2 E82791	conserved hypothet
41	28	93.3	258	2 S45682	acetone-cyanhydrin
42	28	93.3	259	2 S76643	transposase slr051
43	28	93.3	261	2 S71717	transposase sl1171
44	28	93.3	261	2 S73351	transposase sl1171
45	28	93.3	261	2 S75081	transposase slr026

ALIGNMENTS

RESULT 1
N20H1
long neurotoxin 1 - king cobra
N:Alternate names: neurotoxin A
C:Species: Ophiophagus hannah (king cobra)
C>Date: 22-Jun-1981 #sequence_revision 22-Jun-1981 #text_change 23-Aug-1996
C:Accession: A01658
R:Joubert, F.J.
Biochim. Biophys. Acta 317, 85-98, 1973
A:Title: Snake venom toxins. The amino acid sequences of two toxins from Ophiophagus
A:Reference number: A90588; MUID:73231298
A:Accession: A01658
A:Molecule type: protein
A:Residues: 1-73 <JOU>
C:Superfamily: snake toxin
C:Keywords: neurotoxin; venom
F.3-21.14-42.27-31.46-57.58-63/Disulfide bonds: #status predicted

Query Match 96.7%; Score 29; DB 1; Length 73;
Best Local Similarity 80.0%; Pred. No. 33;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
Ib 66 FPMWR 70

RESULT 2
N20H1
translational elongation factor eEF-1 alpha chain - rape (fragment)
N:Alternate names: translation elongation factor EF-1A
C:Species: Brassica napus (rape)
C>Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 31-Mar-2000
C:Accession: T07855
E:Saiz-Vasquez, J.
submitted to the EMBL Data Library, February 1995
A:Reference number: Z16177
A:Accession: T07855
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-113 <SAE>
A:Cross-references: EMBL:U21744; NID:g914912; P1DN:AAA86366.1; P1D:g914913
A:Experimental source: cv. Samourai; 5 days old etiolated seedlings
C:Superfamily: translation elongation factor Tu; translation elongation factor Tu hom
C:Keywords: GTP binding; protein biosynthesis

Query Match 96.7%; Score 29; DB 2; Length 113;
Best Local Similarity 80.0%; Pred. No. 50;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
Db 103 FPMWR 107

RESULT 3

T24380
hypothetical protein T03D8.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24380
R:Morlmore, B.
submitted to the EMBL Data Library, March 1997
A:Reference number: 219884
A:Accession: T24380
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-157 <ML>
A:Cross-references: EMBL:292838; PIDN:CAB07406.1; GSPDB:GN00023; CESP:T03D8.2
A:Experimental source: clone T03D8
C:Genetics:
A:Gene: CESP:T03D8.2
A:Map position: 5
A:Introns: 49/3; 83/1; 125/3

Query Match

Best Local Similarity 96.7%; Score 29; DB 2; Length 157;
Best Local Similarity 80.0%; Pred. No. 68;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
Db 40 FPMWR 44

RESULT 4

D82680
conserved hypothetical protein XF1438 [Imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: D82680
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: D82680
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-245 <SIM>
A:Cross-references: GB:AE003974; GB:AE003849; MID:g9106454; PIDN:AAF84247.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Slipson, A.J.G.; Relbach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briónes, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir, I
as-Neto, E.; Docena, C.; El-Dorry, H.; Facinca, A.P.; Ferreira, A.J.S.
submitted to Genbank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kilaajima, J.P.; Krieger, J.E.; Kurane, E.E.; Latgr
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, F
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, W.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1438

Query Match 96.7%; Score 29; DB 2; Length 245;
Best Local Similarity 80.0%; Pred. No. 1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
Db 197 FPMWR 201

RESULT 5

D85510
probable EC 3.5. amidase-type enzyme yafv [Imported] - Escherichia coli (strain 0157:
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: D85510
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
Iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Diallanta, E.; Potamous, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli 0157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: D85510
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-256 <STO>
A:Cross-references: GB:AE005174; MID:g12512987; PIDN:AAG54544.1; GSPDB:GN00145; UWCP:
A:Experimental source: strain 0157:H7, substrain EDL933
C:Genetics:
A:Gene: yafv

Query Match

Best Local Similarity 96.7%; Score 29; DB 2; Length 256;
Best Local Similarity 80.0%; Pred. No. 1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
Db 244 FPMWR 248

RESULT 6

A29831
heat-labile enterotoxin Ila chain A precursor - Escherichia coli
M:Alternate names: LT-IIa
C:Species: Escherichia coli
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C:Accession: A29831
R:Pickett, C.L.; Weinstein, D.L.; Holmes, R.K.
J. Bacteriol. 169, 5180-5187, 1987
A:Title: Genetics of type Ila heat-labile enterotoxin of Escherichia coli: operon fus
A:Reference number: A91849; MUID:88032841
A:Accession: A29831
A:Molecule type: DNA
A:Residues: 1-259 <PIC>
A:Cross-references: GB:M17894; MID:g146671; PIDN:AAA24093.1; PID:g146672
A:Note: the authors translated the codon TAT for residue 225 as Thr
C:Superfamily: heat-labile enterotoxin chain A
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-259/Product: heat-labile enterotoxin Ila chain A #status predicted <LRA>

Query Match

Best Local Similarity 96.7%; Score 29; DB 2; Length 259;
Best Local Similarity 80.0%; Pred. No. 1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
Db 187 FPMWR 191

RESULT 7

T05454
hypothetical protein F7K2.180 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Sep-1999
C:Accession: T05454
R:Bevan, M.; Medler, H.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schue

submitted to the Protein Sequence Database, November 1998
A:Reference number: Z15416

A:Accession: T05454

A:Molecule type: DNA

A:Residues: 1-273 <BEV>

A:Cross-references: EMBL:AL033545

A:Experimental source: cultivar Columbia; BAC clone F7K2

C:Genetics:

A:Map position: 4

A:Note: F7K2.180

C:Superfamily: Arabidopsis hypothetical protein F7K2.180

Query Match 96.7%; Score 29; DB 2; Length 273;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5

DB 79 FPSMR 83

RESULT 8
C82960
hypothetical protein PA5488 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: G82960

R:Stover, C.K.; Plam, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim,

Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path

A:Reference number: AB2950; MUID:20437337

A:Accession: G82960

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-292 <STO>

A:Cross-references: GB:AE004961; GB:AE004091; NID:g9951814; PIDN:AA08873.1; GSPDB:GN007

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA5488

Query Match 96.7%; Score 29; DB 2; Length 292;
Best Local Similarity 80.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5

DB 231 FPSMR 235

RESULT 9
SS6162
MDCR15 protein - human

C:Species: Homo sapiens (man)

C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999

C:Accession: SS6162

R:Barrella, L.; Loetscher, M.; Tobler, A.; Baggiolini, M.; Moser, B.

Biochem. J. 309, 773-779, 1995

A:Title: Sequence variation of a novel heptahelical leucocyte receptor through alternat

A:Reference number: SS6162; MUID:95366951

A:Accession: SS6162

A:Status: Preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-327 <BAR>

A:Cross-references: EMBL:X68829; NID:g840783; PIDN:CAA8723.1; PID:g840784

C:Superfamily: vertebrate rhodopsin

Query Match 96.7%; Score 29; DB 2; Length 327;
Best Local Similarity 80.0%; Pred. No. 1.4e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5

DB 307 FPSMR 311

RESULT 10

JT0585

minor endoglucanase (EC 3.2.1.-) precursor - Erwinia chrysanthemi

C:Species: Erwinia chrysanthemi

C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 08-Oct-1999

C:Accession: JT0585; JT0586

R:Guiseppi, A.; Aymeric, J.L.; Cami, B.; Barras, F.; Creuzet, N.

Gene 106, 109-114, 1991

A:Title: Sequence analysis of the cellulase-encoding cely gene of Erwinia chrysanthem

A:Reference number: JT0585; MUID:92039050

A:Accession: JT0585

A:Molecule type: DNA

A:Residues: 1-332 <GUI>

A:Cross-references: GB:M74044; NID:g148391; PIDN:AAA24818.1; PID:g148392

A:Experimental source: strain J937

A:Accession: JT0586

A:Molecule type: protein

A:Residues: 24-33 <GUI>

C:Genetics:

A:Gene: cely

A:Keywords: glycosidase; hydrolase

F:1-23/Domain: signal sequence #status predicted <SIG>

F:24-33/Product: minor endoglucanase #status predicted <MIN>

Query Match 96.7%; Score 29; DB 2; Length 332;
Best Local Similarity 80.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5

DB 178 FPSMR 182

RESULT 11
G02913
sperm CD46 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 09-Jun-2000

C:Accession: G02913

R:Hara, T.

submitted to GenBank, March 1996

A:Reference number: H01942

A:Accession: G02913

A:Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-349 <HAR>

A:Cross-references: GB:J04105; NID:g1256700; PIDN:BA12224.1; PID:g1256701

C:Superfamily: herpesvirus complement control protein; complement factor H repeat hom

F:35-94/Domain: complement factor H repeat homology <FH01>

F:99-157/Domain: complement factor H repeat homology <FH02>

F:162-223/Domain: complement factor H repeat homology <FH03>

F:228-283/Domain: complement factor H repeat homology <FH04>

Query Match 96.7%; Score 29; DB 2; Length 349;
Best Local Similarity 80.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5

DB 11 FPSMR 15

RESULT 12
140696

endoglucanase - Cellulomonas uda
 C:Species: Cellulomonas uda
 C:Date: 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 05-Nov-1999
 C:Accession: I40696
 R:Nakamura, K.; Misawa, N.; Kitamura, K.
 J. Biotechnol. 4; 247-254, 1986
 A:Title: Sequence of a cellulase gene of Cellulomonas uda CB4.
 A:Reference number: I40696
 A:Accession: I40696
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1:359 <RES>
 A:Cross-references: GB:M56503; NID:g144430; PIDN:AAA23090.1; PID:g144431

Query Match 96.7% Score 29; DB 2: Length 359;
 Best Local Similarity 80.0%; Pred. No. 1.5e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PPXMR 5
 11 11
 DB 178 FPAWR 182

RESULT 13
 157998
 membrane cofactor protein - human
 C:Species: Homo sapiens (man)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 19-May-2000
 C:Accession: I57998
 R:Cervoni, F.; Fenchel, P.; Akhoundi, C.; Hsi, B.L.; Rossi, B.
 Mol. Reprod. Dev. 34, 107-113, 1993
 A:Title: Characterization of a cDNA clone coding for human testis membrane cofactor protein.
 A:Reference number: I57998; MUID:93119658
 A:Accession: I57998
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-369 <RES>
 A:Cross-references: GB:S51940; NID:g262937; PIDN:AA824802.1; PID:g262938
 C:Superfamily: herpesvirus complement control protein; complement factor H repeat homolog
 F:35-94/Domain: complement factor H repeat homology <FH01>
 F:99-157/Domain: complement factor H repeat homology <FH02>
 F:162-223/Domain: complement factor H repeat homology <FH03>
 F:228-283/Domain: complement factor H repeat homology <FH04>

Query Match 96.7% Score 29; DB 2: Length 369;
 Best Local Similarity 80.0%; Pred. No. 1.5e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PPXMR 5
 11 11
 DB 11 FPSWR 15

RESULT 14
 S26667
 G:protein-coupled receptor BLR1 - human
 C:Species: Homo sapiens (man)
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jul-2000
 C:Accession: S26667
 R:Dobner, T.; Wolf, I.; Emlich, T.; Lipp, M.
 Eur. J. Immunol. 22, 2795-2799, 1992
 A:Title: Differentiation-specific expression of a novel G protein-coupled receptor from A:Reference number: S26667; MUID:93049615
 A:Accession: S26667
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-372 <DOB>
 A:Cross-references: EMBL:X68149; NID:g29459; PIDN:CAA48252.1; PID:g29460
 C:Genetics:
 A:Gene: GDB:BLR1
 A:Cross-references: GDB:136235; OMIM:601613

A:Map position: 15q26.1-15q26.1
 C:Superfamily: vertebrate rhodopsin
 C:Keywords: G protein-coupled receptor; transmembrane protein

Query Match 96.7% Score 29; DB 2: Length 372;
 Best Local Similarity 80.0%; Pred. No. 1.5e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PPXMR 5
 11 11
 DB 352 FPAWR 356

RESULT 15
 154479
 membrane cofactor protein precursor, splice form pm5.1 - human
 N:Alternate names: lymphocyte surface glycoprotein CD46
 C:Species: Homo sapiens (man)
 C:Date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 31-Mar-2000
 C:Accession: 154479
 R:Putcell, D.F.; Russell, S.M.; Deacon, N.J.; Brown, M.A.; Hooker, D.J.; McKenzie, I.
 Immunogenetics 33, 335-344, 1991
 A:Title: Alternatively spliced RNAs encode several isoforms of CD46 (MCP), a regulator of the complement.
 A:Reference number: 154479; MUID:91267562
 A:Accession: 154479
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-377 <RES>
 A:Cross-references: GB:M58050; NID:g180136; PIDN:AAA62833.1; PID:g180137
 C:Genetics:
 A:Gene: GDB:MCP
 A:Cross-references: GDB:120169; OMIM:120920
 A:Map position: 1q32-1q32
 C:Function:
 A:Description: for the factor I-mediated cleavage of the complement convertases
 C:Superfamily: herpesvirus complement control protein; complement factor H repeat homolog
 C:Keywords: alternative splicing; glycoprotein; transmembrane protein
 F:35-94/Domain: complement factor H repeat homology <FH01>
 F:99-157/Domain: complement factor H repeat homology <FH02>
 F:162-223/Domain: complement factor H repeat homology <FH03>
 F:228-283/Domain: complement factor H repeat homology <FH04>
 F:329-351/Domain: transmembrane #status predicted <TM>

Query Match 96.7% Score 29; DB 2: Length 377;
 Best Local Similarity 80.0%; Pred. No. 1.6e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PPXMR 5
 11 11
 DB 11 FPAWR 15

Search completed: February 27, 2002, 11:45:22
 Job time: 701 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:42:47 ; Search time 78.39 Seconds
(without alignments)
2.339 Million cell updates/sec

Title: US-09-446-109A-13

Perfect score: 30
Sequence: 1 FPXMR 5

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	29	96.7	73 1 NXLL_OPHHA	P01387 ophiophagus
2	29	96.7	165 1 DSRB_VIBAL	O56578 vibrio algi
3	29	96.7	259 1 E2MA_ECOLI	P1810 escherichia
4	29	96.7	332 1 GUNY_ERWCH	P27032 erwina chr
5	29	96.7	359 1 GUN_CELUD	P18336 cellulomona
6	29	96.7	372 1 CCR5_HUMAN	P13302 homo sapien
7	29	96.7	377 1 MCP_HUMAN	P15529 homo sapien
8	29	96.7	420 1 Z229_HUMAN	O91477 homo sapien
9	28	93.3	143 1 H516_CAEEL	P06581 caenorhabdi
10	28	93.3	143 1 H517_CAEEL	P05513 caenorhabdi
11	28	93.3	257 1 HNL_MANES	P52705 manihot esc
12	28	93.3	271 1 PABC_VIBHA	O56693 vibrio harv
13	28	93.3	271 1 PPMK_BACST	P58055 bacillus st
14	28	93.3	286 1 PA1_KLEPN	P37446 klebsiella
15	28	93.3	289 1 PA1_ECOLI	P00631 escherichia
16	28	93.3	289 1 PA1_PROVU	P37447 proteus vul
17	28	93.3	289 1 PA1_SALTY	P37442 salmonella
18	28	93.3	298 1 CC28_YEAST	P00546 saccharomyc
19	28	93.3	332 1 CC2_GAEEL	P34556 caenorhabdi
20	28	93.3	374 1 CCR5_MOUSE	O04683 mus musculu
21	28	93.3	374 1 CCR5_RAT	P34997 rattus norv
22	28	93.3	430 1 KDTA_CHILMU	O9PK15 chlamydia m
23	28	93.3	431 1 KDTA_CHLTR	O57440 chlamydia t
24	28	93.3	433 1 ANM2_HUMAN	P55345 homo sapien
25	28	93.3	437 1 KDTA_CHLPN	O46222 chlamydia p
26	28	93.3	448 1 Y347_CHLPN	O94816 chlamydia p
27	28	93.3	459 1 RBL2_THIDE	O60028 thiodacilli
28	28	93.3	463 1 RBL3_HYDMR	O59462 hydrogenvi
29	28	93.3	470 1 LIP2_MYOCO	O64424 myocastor c
30	28	93.3	482 1 CATR_ONCVC	O27710 onchocora
31	28	93.3	490 1 CE05_ECOLI	O47500 escherichia
32	28	93.3	490 1 CE10_ECOLI	O47125 escherichia
33	28	93.3	492 1 CAT2_CUCPE	P48351 cucurbita p

ALIGNMENTS

RESULT	ID	STANDARD	PRT	73 AA.
1	NXLL_OPHHA			
AC	P01387			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	21-JUL-1986 (Rel. 01, last sequence update)			
DT	20-AUG-2001 (Rel. 40, last annotation update)			
DE	LONG NEUROTOXIN 1 (NEUROTOXIN A)			
OS	Ophiophagus hannah (King cobra) (Naja hannah).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Lepidodactylus; Squamata; Scleroglossa; Serpentes; Colubroidae;			
OC	Elapidae; Elapinae; Ophiophagus.			
OX	NCBI_Taxid=8665;			
RN	(1)			
RP	SEQUENCE.			
RC	TISSUE=Venom.			
RC	MEDLINE=7321298; PubMed=4198767;			
HA	Joubert F.J.;			
KT	"Snake venom toxins the amino acid sequences of two toxins from			
KT	Biochim. Biophys. Acta 317:85-98(1973).			
KL	Ophiophagus hannah (King cobra) venom."			
CC	-1- SUBCELLULAR LOCATION: SECRETED.			
CC	-1- TISSUE SPECIFICITY: EXPRESSED BY VENOM GLAND.			
CC	-1- MISCELLANEOUS: LD(50) IS 0.3 MG/KG BY SUBCUTANEOUS INJECTION.			
CC	-1- SIMILARITY: BELONGS TO THE SNAKE TOXIN FAMILY.			
DR	PIR: A01658; N2OHL.			
DR	HSSP: P01386; ITXB.			
DR	Interpro: IPR003571; Snake_toxin.			
DR	Pfam: PF00087; toxin; 1.			
DR	ProDom: PD000206; Snake_toxin; 1.			
DR	PROSITE: PS00272; SNAKE_TOXIN; 1.			
KW	Venom; Neurotoxin; Multigene family.			
FT	DISULFID 3 21			
FT	DISULFID 14 42			
FT	DISULFID 27 31			
FT	DISULFID 46 57			
FT	DISULFID 58 63			
FT	SEQUENCE 73 AA: 8106 MW: 1AC17E91E1C54F7 CRC64:			
QY	Query Match	96.7%: Score 29; DB 1; Length 73;		
DB	Best Local Similarity	80.0%: Pred. NO. 17;		
DB	Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
QY	1 FPXMR 5			
DB	66 FETWR 70			
RESULT	2			
ID	DSBB_VIBAL	STANDARD:	PRT:	165 AA.
AC	O56578;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, last sequence update)			
DT	20-AUG-2001 (Rel. 40, last annotation update)			

```

DE DISULFIDE BOND FORMATION PROTEIN B (DISULFIDE OXIDOREDUCTASE)
DE (FRAGMENT).
CN DSB.
OS Vibrio alginolyticus.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=663;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=BJ38-2;
RA Nakamura T., Enomoto H., Unemoto T.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR DISULFIDE BOND FORMATION IN SOME
CC PERIPLASMIC PROTEINS. ACTS BY OXIDIZING THE DSB. PROTEIN (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE DSB. FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D83728; BAA12087.1;
DR InterPro: IPR003752; DSB.
DR Pfam: PF02600; DSB.1
KW Oxidoreductase; Redox-active center; Electron transport; Chaperone;
KW Transmembrane; Inner membrane.
FT DOMAIN 3 16 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 17 33 POTENTIAL.
FT DOMAIN 34 51 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 52 67 POTENTIAL.
FT DOMAIN 68 74 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 75 92 POTENTIAL.
FT DOMAIN 93 147 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 148 165 POTENTIAL.
FT DISULFID 43 46 REDOX-ACTIVE (BY SIMILARITY).
FT DISULFID 107 133 REDOX-ACTIVE (BY SIMILARITY).
FT NON_TER 165 165
SQ SEQUENCE 165 AA: 18745 MW: 043696DF5AA3272D CRC64:

Query Match 96.7%: Score 29; DB 1; Length 165;
Best Local Similarity 80.0%: Pred. No. 37;
Matches 4: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPMWR 5
DB 113 FPMWR 117

RESULT 3
E2AA_ECOLI STANDARD: PRT: 259 AA.
ID E2AA_ECOLI STANDARD: PRT: 259 AA.
AC P13810;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HEAT-LABILE ENTEROTOXIN IIA, A CHAIN PRECURSOR (LT-IIA).
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN 11
RP SEQUENCE FROM N.A.
RA MEDLINE=88032841; PubMed=2822667;
RA Pickert C.L., Weinstein D.L., Holmes R.K.;
RT "Genetics of type IIA heat-labile enterotoxin of Escherichia coli:
RT operon fusions, nucleotide sequence, and hybridization studies.";
RL J. Bacteriol. 169:5180-5187(1987).

```

```

CC -1- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY
CC THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.
CC -1- SUBUNIT: HETEROHEXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M17894; AAA24093.1;
DR PIR: A29831; A29831.
DR HSSP: P43528; ITII.
DR InterPro: IPR001144; Enterotoxin_A.
DR Pfam: PF01375; Enterotoxin_A.1.
DR PRINTS: PR00771; ENTEROTOXIN.
KW Enterotoxin; Signal.
FT SIGNAL 1 18
FT CHAIN 19 259 HEAT-LABILE ENTEROTOXIN IIA, A CHAIN.
FT DISULFID 203 215 BY SIMILARITY.
FT ACT_SITE 128 128
SQ SEQUENCE 259 AA: 29242 MW: 996F311A32CABEA CRC64:

Query Match 96.7%: Score 29; DB 1; Length 259;
Best Local Similarity 80.0%: Pred. No. 56;
Matches 4: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPMWR 5
DB 187 FPMWR 191

RESULT 4
ID GUNY_ERWCH STANDARD: PRT: 332 AA.
AC P27032;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE MINOR ENDOLUCINASE Y PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE
DE Y) (CELLULOSE Y) (EGY).
GN CELY.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=766;
RN 11
RP SEQUENCE FROM N.A. AND SEQUENCE OF 24-33.
RC STRAIN=3937;
RX MEDLINE=92039050; PubMed=1937031;
RX Giuseppe A., Aymeric J.-L., Cami B., Barras F., Creuzet N.;
RX "Sequence analysis of the cellulase-encoding cely gene of Erwinia
RX chrysanthemi: a possible case of interspecies gene transfer.";
RL Gene 106:109-114(1991).
CC -1- FUNCTION: REPRESENTS ONLY 3% OF THE GLOBAL CELLULOSE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: ENDOLUCINASE OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY D (FAMILY 8 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M74044; AAA24818.1;
DR PIR: JT0585; JT0585.

```

DR InterPro: IPR002037; Glyco_hydro_8.
 DR Pfam: PF01270; Glyco_hydro_8; 1.
 DR PRINTS: PR00735; GLHYDRLASE8.
 DR PROSITE: PS00812; GLYCOSYL_HYDROL_F8; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 23
 FT CHAIN 1 23
 FT ACT_SITE 53 53 MINOR ENDOGLUCANASE Y.
 FT ACT_SITE 110 110 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 110 110 NUCLEOPHILE (POTENTIAL).
 SO SEQUENCE 332 AA: 37592 MW: 65A760A2B8227079 CRC64;

Query Match: 96.7%; Score 29; DB 1; Length 332;
 Best Local Similarity 80.0%; Pred. No. 71;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FPMXR 5
 Db 178 FPMXR 182

RESULT 5
 GUN_CELUD STANDARD; PRT: 359 AA.
 AC P18336;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
 OS (CELLULOMONAS).
 OC Cellulomonas uda.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Micrococcales; Cellulomonadaceae; Cellulomonas.
 OX NCBI_TaxID=1714;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-37.
 RC STRAIN=C84;
 RA Nakamura K., Misawa N., Kitamura K.;
 RT "Sequence of a cellulase gene of Cellulomonas uda C84."
 RL J. Biochem. 4:247-254(1986).
 CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
 CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
 CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
 CC (2) EXOCELLULOBIODHOLASES THAT CUT THE DISACCHARIDE CELLULOSE
 CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
 CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER
 CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY D (FAMILY 8 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M36503; AAA23090.1; -
 CC InterPro: IPR002037; Glyco_hydro_8.
 DR Pfam: PF01270; Glyco_hydro_8; 1.
 DR PROSITE: PS00812; GLYCOSYL_HYDROL_F8; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 23
 FT CHAIN 1 23
 FT ACT_SITE 53 53 ENDOGLUCANASE.
 FT ACT_SITE 110 110 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 110 110 NUCLEOPHILE (POTENTIAL).
 SO SEQUENCE 359 AA: 40690 MW: 0445D571B683148 CRC64;

Query Match 96.7%; Score 29; DB 1; Length 359;

Best Local Similarity 80.0%; Pred. No. 76;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 FPMXR 5
 Db 178 FPMXR 182

RESULT 6
 CCR5_HUMAN STANDARD; PRT: 372 AA.
 ID CCR5_HUMAN
 AC P32302; Q14811;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE C-X-C CHEMOKINE RECEPTOR TYPE 5 (CXCR-5) (CXCR-5) (BURKITT'S LYMPHOMA
 DE RECEPTOR 1) (MONOCYTE-DERIVED RECEPTOR 15) (MDR15).
 GN BURL OR CXCR5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG FORM).
 RC TISSUE=Lymphocytes;
 RX MEDLINE=93049615; PubMed=1425907;
 RA Dobner T., Wolf I., Emrich T., Lipp M.;
 RT "Differentiation-specific expression of a novel G protein-coupled
 RT receptor from Burkitt's lymphoma."
 RL Eur. J. Immunol. 22:2795-2799(1992).
 RN [2]
 RP SEQUENCE FROM N.A. (SHORT FORM).
 RC TISSUE=Blood;
 RX MEDLINE=95366951; PubMed=7639692;
 RA Battella L., Loetscher M., Tobler A., Baggiolini M., Moser B.;
 RT "Sequence variation of a novel heptahelical leucocyte receptor
 RT through alternative transcript formation."
 RL Biochem. J. 309:773-779(1995).
 RN [3]
 RP LIGAND BINDING.
 RX MEDLINE=98130629; PubMed=9463416;
 RA Legler D.F., Loetscher M., Stuber Roos R., Clark-Lewis I.,
 RA Baggiolini M., Moser B.;
 RT "B cell-attracting chemokine 1, a human CXC chemokine expressed in
 RT lymphoid tissues, selectively attracts B lymphocytes via
 RT Bln/CXCR5."
 RL J. Exp. Med. 187:655-660(1998).
 CC -1- FUNCTION: CYTOKINE RECEPTOR THAT BINDS TO BLC. BURL EXERTS
 CC POSSIBLY A REGULATORY FUNCTION IN BURKITT LYMPHOMA (BL)
 CC LYMPHOMAGENESIS AND/OR B-CELL DIFFERENTIATION. IT IS A POTENTIAL
 CC CANDIDATE FOR CELL-CELL INTERACTION, AND ACTIVATION OF MATURE B-
 CC LYMPHOCYTES IN LYMPHATIC TISSUES.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSION IN MATURE B-CELLS AND BURKITT
 CC LYMPHOMA CELLS.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X68149; CAA48252.1; -
 CC EMBL: X68829; CAA48723.1; -
 DR PIR: S26667; S26667.
 DR GCRdb: GCR_0453; -
 DR GCRdb: GCR_2072; -
 DR GCRdb: GCR_2612; -

MM: 601613: -
 DR InterPro: IPR000276; GPCR_Rhodopsn.
 DR Pfam: PF00001; 7tm_1.1
 DR PRINTS: PR00237; GPCR_Rhodopsn.
 DR PRINTS: PR00564; BURKITSLEYMR.
 DR PROSITE: PS00237; G-PROTEIN_RECPT_FL_1; 1.
 DR PROSITE: PS02624; G-PROTEIN_RECPT_FL_2; 1.
 DR G-protein coupled receptor; Transmembrane; Glycoprotein; B-cell;
 KW Alternative splicing.
 FT DOMAIN 1 55
 FT TRANSMEM 56 76
 FT DOMAIN 77 88
 FT TRANSMEM 89 109
 FT DOMAIN 110 124
 FT TRANSMEM 125 145
 FT DOMAIN 146 167
 FT TRANSMEM 168 188
 FT DOMAIN 189 219
 FT TRANSMEM 220 240
 FT DOMAIN 241 259
 FT TRANSMEM 260 280
 FT DOMAIN 281 304
 FT TRANSMEM 305 325
 FT DOMAIN 326 372
 FT CARBOHYD 28 28
 FT CARBOHYD 196 196
 FT DISULFID 122 202
 FT VAARSPLIC 1 45
 FT CONFLICT 344 344
 FT SEQUENCE 372 AA; 41955 MW; 6D6F84C83942ACFC CRC64;

Query Match 96.7%; Score 29; DB 1; Length 372;
 Best Local Similarity 80.0%; Pred. No. 79;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
 DB 352 FPSWR 356

RESULT 7
 MCP_HUMAN
 ID MCP_HUMAN STANDARD: PRT; 377 AA.
 AC P15529;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE MEMBRANE COFACTOR PROTEIN PRECURSOR (CD46 ANTIGEN) (TROPHOBLAST
 DE LEUCOCYTE COMMON ANTIGEN) (TLX).
 GN MCP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN 11
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 35-58.
 RX MEDLINE=88286080; PubMed=3260937;
 RA Lublin D.M., Liszewski M.K., Post T.W., Arce M.A., Le Beau M.M.,
 RA Rebenitsch M.B., Lemons R.S., Seya T., Atkinson J.P.;
 RT "Molecular cloning and chromosomal localization of human membrane
 RT cofactor protein (MCP). Evidence for inclusion in the multigene
 RT family of complement-regulatory proteins.";
 RL J. Exp. Med. 168:181-194(1988).
 RN 12
 RP SEQUENCE FROM N.A.
 RC TISSUE-Testis.
 RX MEDLINE=93119658; PubMed=8418811;
 RA Cervoni F., Fenichel P., Akhoundi C., Hsi B.L., Rossi B.;
 RT "Characterization of a cDNA clone coding for human testis membrane
 RT cofactor protein (MCP, CD46)."
 RL Mol. Reprod. Dev. 34:107-113(1993).
 RN 13

RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
 RX MEDLINE=91267562; PubMed=2050389;
 RA Purcell D.F., Russell S.M., Deacon N.J., Brown M.A., Hooker D.J.,
 RA McKenzie I.F.;
 RT "Alternatively spliced RNAs encode several isoforms of CD46 (MCP), a
 RT regulator of complement activation.";
 RL Immunogenetics 33:335-344(1991).
 RN 14
 RP SEQUENCE OF 1-34 FROM N.A.
 RX MEDLINE=94014356; PubMed=7691939;
 RA Cui W., Hourcade D., Post T., Greenlund A.C., Atkinson J.P.,
 RA Kumar V.;
 RT "Characterization of the promoter region of the membrane cofactor
 RT protein (CD46) gene of the human complement system and comparison to
 RT a membrane cofactor protein-like genetic element.";
 RL J. Immunol. 151:4137-4146(1993).
 RN 15
 RP ALTERNATIVE SPLICING.
 RX MEDLINE=92289809; PubMed=1601037;
 RA Russell S.M., Sparrow R.L., McKenzie I.F.C., Purcell D.F.J.;
 RT "Tissue-specific and allelic expression of the complement regulator
 RT CD46 is controlled by alternative splicing.";
 RL Eur. J. Immunol. 22:1513-1518(1992).
 CC -1- FUNCTION: MCP IS AN IMPORTANT MEMBRANE PROTEIN FOR PROTECTING HOST
 CC CELLS FROM DAMAGE BY COMPLEMENT. MCP POSSESSES COFACTOR ACTIVITY
 CC FOR INACTIVATION OF C3B AND C4B BY SERUM FACTOR I. THIS WOULD
 CC ALLOW MCP TO REGULATE THE COMPLEMENT SYSTEM AT THE LEVEL OF THE C3
 CC CONVERTASES. MAY BE INVOLVED IN THE FUSION OF THE SPERMATOZOA WITH
 CC THE OOCYTE DURING FERTILIZATION. MAY ACT BY PREVENTING COMPLEMENT
 CC ACTIVATION AT THE FETO-MATERNAL INTERFACE ON THE
 CC STINCT TROPHOBLAST LAYER OF PLACENTA.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 14 FORMS ARE PRODUCED BY
 CC ALTERNATIVE SPLICING. THE FORM SHOWN IS THE TRANSCRIPT A. THE
 CC GAMMA ISOFORM (TRANSCRIPTS A AND B) IS PREFERENTIALLY EXPRESSED IN
 CC BEV-B CELLS AND LEUKEMIC CELLS. THE 66 KDA ALPHA ISOFORM AND THE
 CC 56 KDA BETA ISOFORM FOUND IN ALL TISSUES EXCEPT SPERM CORRESPOND
 CC TO THE TRANSCRIPTS C AND D, AND THE TRANSCRIPTS E AND F
 CC RESPECTIVELY. THE EXON 9 IS SPECIFICALLY DELETED IN SOME
 CC PLACENTAE. THE EPSILON ISOFORM CORRESPONDS TO THE TRANSCRIPTS I
 CC AND J. SPERMATOZOA DELTA ISOFORM (TRANSCRIPT N) DELETE BOTH EXONS
 CC 12 AND 13. ALL TISSUES DIFFERENTIALLY SPLICE EXON 13.
 CC -1- TISSUE SPECIFICITY: EXPRESSED BY ALMOST ALL CELLS EXCEPT
 CC ERYTHROCYTES AND SOME BONE MARROW CELLS.
 CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED IN THE SER/THR-RICH DOMAIN
 CC (PROBABLY).
 CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
 CC DATABASE: NAME-PROM: NOTE-CD guide CD46 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd46.htm".
 CC -1- THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL, outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Y00651; CAA68675.1; -
 CC EMBL: S51940; AAB24802.1; -
 CC EMBL: M56050; AAB62833.1; -
 CC EMBL: A16585; CAA01400.1; -
 CC EMBL: S65879; AAD13968.1; -
 CC PIR: S01896; S01896.
 CC HSSP: P10998; LVVC.
 CC MTM: I20920; -
 DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF00084; sushi_4.
 DR SMART: SM00032; CCP; 4.
 KW Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
 KW Sushi; Alternative splicing.
 FT SIGNAL 1 34
 FT CHAIN 35 377 MEMBRANE COFACTOR PROTEIN.

```

FT DOMAIN 35 328 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 329 351 POTENTIAL.
FT DOMAIN 352 377 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 353 95 SUSHI 1.
FT DOMAIN 354 158 SUSHI 2.
FT DOMAIN 161 224 SUSHI 3.
FT DOMAIN 227 284 SUSHI 4.
FT DOMAIN 287 311 SER/THR-RICH.
FT DISULFID 35 80 BY SIMILARITY.
FT DISULFID 64 94 BY SIMILARITY.
FT DISULFID 99 141 BY SIMILARITY.
FT DISULFID 127 157 BY SIMILARITY.
FT DISULFID 210 223 BY SIMILARITY.
FT DISULFID 191 223 BY SIMILARITY.
FT DISULFID 228 270 BY SIMILARITY.
FT DISULFID 236 283 BY SIMILARITY.
FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 273 273 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPPLIC 286 300 MISSING (IN A SECOND FORM).
FT VARSPPLIC 340 352 MISSING (IN ISOFORM M).
FT VARSPPLIC 353 361 YLDKRRKKG -> DTRGGRKKGKQWELNMLPLRLNPLQ
FT VARSPPLIC 362 377 QSRKAE (IN ISOFORM M).
FT VARSPPLIC 362 377 TYLDETHREKFTSL -> KADGAEVATYQTKSTTPAEQ
FT VARSPPLIC 362 377 RG (IN ISOFORM B/D/F/H/J/L).
FT VARSPPLIC 340 361 MISSING (IN ISOFORM M AND ISOFORM N).
FT VARSPPLIC 339 339 MISSING (IN ISOFORM N).
FT VARSPPLIC 339 339 I -> IGKQWELNMLPLRLNPLQOQSRKAE (IN
FT SEQUENCE 377 AA: 42247 MW: 2CAG6F1752570B57 CRC64:
SQ
Query Match 96.7% Score 29: DB 1: Length 377:
Best Local Similarity 80.0% Pred. No. 80:
Matches 4: Conservative 0: Mismatches 1: Indels 0: Gaps 0:
QY 1 FPXMR 5
DB 11 FPMR 15

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AF192979; AAF07964.1; -
CC EMBL: AC084239; AAG23970.1; -
CC InterPro: IPR001909; KRAB.
CC InterPro: IPR000822; ZnF-C2H2.
CC Pfam: PF01352; KRAB.1.
CC Pfam: PF00096; zf-C2H2.2.
CC PRINTS: PR00048; ZINC_FINGER.
CC SMART: SM00349; KRAB.1.
CC SMART: SM00355; ZnF_C2H2.2.
CC PROSITE: PS00805; KRAB.1.
CC PROSITE: PS00028; ZINC_FINGER_C2H2.1; 2.
CC PROSITE: PS00157; ZINC_FINGER_C2H2.2; 3.
CC Transcription regulation: DNA-binding, Zinc-finger: Metal-binding;
CC Nuclear Protein: Repeat.
CC DOMAIN 34 108 KRAB.
CC FT DOMAIN 349 >420 ZINC_FINGERS.
CC FT ZN_FING 349 371 C2H2-TYPE.
CC FT ZN_FING 377 399 C2H2-TYPE.
CC FT ZN_FING 405 >420 C2H2-TYPE.
CC FT NON_TER 420 420
CC SEQUENCE 420 AA: 48022 MW: FA4138BA4A4A2A14 CRC64:
SQ
Query Match 96.7% Score 29: DB 1: Length 420:
Best Local Similarity 80.0% Pred. No. 89:
Matches 4: Conservative 0: Mismatches 1: Indels 0: Gaps 0:
QY 1 FPXMR 5
DB 173 FPMR 177

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: M14334; AAA28070.1; ALT_SEQ.
 DR EMBL: X01577; CAA25732.1; -
 DR PIR: A38884; HHK41.
 DR PIR: A25199; A25199.
 DR InterPro: IPR002068; Crystallin_HSP20.
 DR Pfam: PF00031; HSP20; 1.
 DR PROSITE: PS01031; HSP20; 1.
 DR Heat shock: Multigene family.
 SW SEQUENCE 143 AA; 16252 MW; C1D0F59D26E36C24 CRC64;

Query Match 93.3%; Score 28; DB 1; Length 143;
 Best Local Similarity 80.0%; Pred. No. 50;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPYWR 5
 Db 27 FPYWR 31

RESULT 10
 HS17-CAEEL STANDARD: PRT: 143 AA.
 AC P02513;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE HEAT SHOCK PROTEIN HSP16-48.
 GN HSP16-48.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 NCBI_TaxID=6239;
 RX MEDLINE=85295957; PubMed=4033652;
 RA Russnak R.H., Candido E.P.M.;
 RT "Locus encoding a family of small heat shock genes in Caenorhabditis
 RT elegans: two genes duplicated to form a 3.8-kilobase inverted
 RT repeat";
 RL Mol. Cell. Biol. 5:1268-1278(1985).
 RN (2)
 RP SEQUENCE OF 9-143 FROM N.A.
 RX MEDLINE=8320736; PubMed=6190129;
 RA Russnak R.H., Jones D., Candido E.P.M.;
 RT "Cloning and analysis of cDNA sequences coding for two 16 kilodalton
 RT heat shock proteins (hsps) in Caenorhabditis elegans: homology with
 RT the small hsps of Drosophila";
 RL Nucleic Acids Res. 11:3187-3205(1983).
 CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: K03273; AAA28069.1; -
 DR EMBL: X01576; CAA25731.1; -
 DR EMBL: K01863; AAA28064.1; -
 DR PIR: A02916; HHK48.
 DR PIR: A24289; A24289.
 DR InterPro: IPR002068; Crystallin_HSP20.

CR Pfam: PF00011; HSP20; 1.
 DR PROSITE: PS01031; HSP20; 1.
 SW Heat shock: Multigene family.
 SW SEQUENCE 143 AA; 16299 MW; 0D5596DFE5B3118 CRC64;

Query Match 93.3%; Score 28; DB 1; Length 143;
 Best Local Similarity 80.0%; Pred. No. 50;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPYWR 5
 Db 27 FPYWR 31

RESULT 11
 HNL_MANES STANDARD: PRT: 257 AA.
 AC P52705;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE (S)-ACETONE-CYANOHYDRIN LYASE (EC 4.1.2.39) ((S)-HYDROXYNITRILE LYASE)
 DE ((S)-HYDROXYNITRILASE) (OXYNITRILASE).
 GN HNL.
 OS Manihot esculenta (Cassava) (Manioc).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Malpighiales; Euphorbiaceae; Manihot.
 NCBI_TaxID=3983;
 RX MEDLINE=94263231; PubMed=8203915;
 RA Hughes J., Carvalho F.J.P.D.C., Hughes M.A.;
 RT "Purification, characterization, and cloning of alpha-hydroxynitrile
 RT lyase from cassava (Manihot esculenta Crantz).";
 RL Arch. Biochem. Biophys. 311:496-502(1994).
 CC -1- FUNCTION: INVOLVED IN CYANOGENESIS. THE RELEASE OF HCN FROM
 CC INJURED TISSUES. DECOMPOSES A VARIETIES OF (R) OR (S) CYANOHYDRINS
 CC INTO HCN AND THE CORRESPONDING ALDEHYDES AND KETONES. THE NATURAL
 CC SUBSTRATE OF THIS ENZYME IS (S)-ACETONE CYANOHYDRIN.
 CC -1- CATALYTIC ACTIVITY: (S)-2-HYDROXYISOBUTYRONITRILE + CYANIDE +
 CC ACETONE.
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- SIMILARITY: DISTANT RELATIONSHIP WITH TYPE-B
 CC CARBOXYL ESTERASE/LIPASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: Z29091; CAA82334.1; -
 DR HSP: P52704; 1YAS
 DR InterPro: IPR000073; Abhydrolase.
 DR InterPro: IPR000379; Est_Lip_cholesterol_acylase.
 DR Pfam: PF00561; abhydrolase; 1.
 KW Lyase.
 FT INIT.MET 0 0
 FT ACT_SITE 79 79 BY SIMILARITY.
 FT ACT_SITE 207 207 BY SIMILARITY.
 FT ACT_SITE 235 235 BY SIMILARITY.
 FT SEQUENCE 257 AA; 29240 MW; 98B3E160ACB338C5 CRC64;

Query Match 93.3%; Score 28; DB 1; Length 257;
 Best Local Similarity 80.0%; Pred. No. 87;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
11 11
Db 124 FPDWR 128

RESULT 12

PABC_VIBHA STANDARD: PRT: 271 AA.
PABC_VIBHA
AC 056693;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE 4-AMINO-4-DEOXYCHORISMATE LYASE (EC 4.-.-.-) (ADC LYASE) (ADCL).
GN PABC.
OS Vibrio harveyi.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=669;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN-B392;
RX MEDLINE=96134997; PubMed=8550484;
RA Shen Z., Byers D.M.;
RT "Isolation of Vibrio harveyi acyl carrier protein and the fabG, acpP,
and fabF genes involved in fatty acid biosynthesis."
RL J. Bacteriol. 178:571-573(1996).
CC -1- FUNCTION: CONVERTS 4-AMINO-4-DEOXYCHORISMATE INTO 4-AMINOBENZOATE
(PABA) AND PYRUVATE.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- PATHWAY: SECOND STEP IN FOLATE BIOSYNTHESIS PATHWAY.
CC -1- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT
AMINOTRANSFERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U39441; AAC43592.1; -;
DR InterPro: IPR001544; AminoTran_4.
DR Pfam: PF01063; aminoTran_4; 1.
DR ProDom: PD001961; AminoTran_4; 1.
DR PROSITE: PS00770; AA_TRANSFER_CLASS_4; 1.
KM lysase; Pyridoxal phosphate; Folate biosynthesis.
FT PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT 140 140
SQ SEQUENCE 271 AA: 29558 MW: 3A5BEB9F398D1CD CRC64;

Query Match 93.38; Score 28; DB 1; Length 271;
Best Local Similarity 80.08; Pred. No. 91;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
11 11
Db 57 FPDWR 61

RESULT 13

PPNK_BACST STANDARD: PRT: 271 AA.
AC P58055;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE INORGANIC POLYPHOSPHATE/ATP-NAD KINASE (EC 2.7.1.23)
DE (POLY(P)/ATP NAD KINASE).
GN PPNK OR NADK.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Geobacillus.

OX NCBI_TaxID=1422;

RN 111
RP SEQUENCE FROM N.A.
RA Ohshima T., Sakuraba H.;
RT "Inorganic polyphosphate/ATP-NAD kinase of Bacillus
stearothermophilus."
KL Submitted (Feb-2001) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: CATALYZES THE PHOSPHORYLATION OF NAD TO NADP. UTILIZES
CC ATP AND OTHER NUCLEOSIDE TRIPHOSPHATES AS WELL AS INORGANIC
CC POLYPHOSPHATE AS A SOURCE OF PHOSPHORUS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + NAD(+) -> ADP + NADP(+).
CC -1- COFACTOR: REQUIRES DIVALENT METAL IONS FOR ACTIVITY (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NAD KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AB055961; BAB32727.1; -;
DR Transferase; Kinase; NAD; NADP.
KM SEQUENCE 271 AA: 30517 MW: 9E0FB427BC76662A CRC64;

Query Match 93.38; Score 28; DB 1; Length 271;
Best Local Similarity 80.08; Pred. No. 91;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
11 11
Db 258 FPDWR 262

RESULT 14

PAL_KLEPN STANDARD: PRT: 286 AA.
AC P37446;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PHOSPHOLIPASE A1 PRECURSOR (EC 3.1.1.32) (DETERGENT-RESISTANT
DE PHOSPHOLIPASE A) (DR-PHOSPHOLIPASE A) (PHOSPHATIDYLCHOLINE 1-
DE ACYLHYDROLASE) (OUTER MEMBRANE PHOSPHOLIPASE A) (OM PLA).
GN PLDA.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=573;
RN 111
RP SEQUENCE FROM N.A.
RX MEDLINE=94131966; PubMed=8300539;
RA Brok R.G.P.M., Brinkman E., van Boxtel R., Bekkers A.C.A.P.,
PA Verheij H.M., Tommassen J.;
RT "Molecular characterization of enterobacterial plda genes encoding
KT outer membrane phospholipase A."
KL J. Bacteriol. 176:861-870(1994).
CC -1- FUNCTION: HYDROLYSIS OF PHOSPHATIDYLCHOLINE WITH PHOSPHOLIPASE
CC A2 (EC 3.1.1.4) AND PHOSPHOLIPASE A1 (EC 3.1.1.32) ACTIVITIES.
CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O -> 1-ACYLGLYCERO-
CC PHOSPHOCHOLINE + A FATTY ACID ANION.
CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O -> 2-ACYLGLYCERO-
CC PHOSPHOCHOLINE + A FATTY ACID ANION.
CC -1- COFACTOR: REQUIRES CALCIUM IONS FOR ACTIVITY.
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE; ONE OF THE VERY FEW ENZYMES
CC LOCATED THERE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL X76901: CAA54223.1; -
 DR PIR: B36971; B36971.
 DR PIR: S40129; S40129.
 DR InterPro: IPR003187; PLAI.
 DR Pfam: PF02253; PLAI: 1.
 KM Hydrolyase; Lipid degradation: Outer membrane: Signal: Calcium.
 FT SIGNAL: 1 20 BY SIMILARITY.
 FT CHAIN: 21 286 PHOSPHOLIPASE A1.
 FT ACT_SITE: 161 161 BY SIMILARITY.
 SQ SEQUENCE 286 AA: 32544 MW: 3E39F863085108A3 CRC64:

Query Match 93.38; Score 28; DB 1; Length 286;
 Best Local Similarity 80.0%; Pred. No. 96;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 Db 92 FPLMR 96

RESULT 15
 ID PAL-ECOLI STANDARD: PRT: 289 AA.
 AC P00531;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PHOSPHOLIPASE A1 PRECURSOR (EC 3.1.1.32) (DEREGENT-RESISTANT
 DE PHOSPHOLIPASE A) (DR-PHOSPHOLIPASE A) (PHOSPHATIDYLCHOLINE 1-
 DE ACYLHYDROLASE) (OUTER MEMBRANE PHOSPHOLIPASE A) (OM PLA).
 GN PLDA OR B3821 OR Z5342 OR ECS4751.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562, 83334;
 RN 11
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85157492; PubMed=6397464;
 RA Homma H., Kobayashi T., Chida N., Karasawa K., Mizushima H., Kudo I.,
 RA Inoue K., Ikeda H., Sekiguchi M., Nojima S.;
 RT "The DNA sequence encoding plda gene, the structural gene for
 RT detergent-resistant phospholipase A of E. coli.";
 RL J. Biochem. 96:1655-1664(1984).
 RN 12
 RP SEQUENCE FROM N.A.
 RX STRAIN-K12 / M01655;
 RX MEDLINE=92358234; PubMed=1379743;
 RA Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;
 RT "Analysis of the Escherichia coli genome: DNA sequence of the region
 RT from 84.5 to 86.5 minutes.";
 RL Science 257:771-778(1992).
 RN 13
 RP REVISION TO 14-15.
 RX STRAIN-K12 / M01655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN 14
 RP SEQUENCE FROM N.A.
 RX STRAIN-O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postell G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dinalanta E.T., Potamouis K.,
 RA Apodaca J., Mantharman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch K.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533(2001).
 RN 15
 RP SEQUENCE FROM N.A.
 RX STRAIN-O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shingawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 RN 16
 RP SEQUENCE OF 30-289 FROM N.A.
 RX STRAIN-K12;
 RX MEDLINE=85003590; PubMed=6383820;
 RA de Geus P., Verheij H.M., Riegman N.H., Hoeksma W.P.M., de Haas G.H.;
 RT "The pro- and mature forms of the E. coli K-12 outer membrane
 RT phospholipase A are identical.";
 RL EMBO J. 3:1799-1802(1984).
 RN 17
 RP SEQUENCE OF 174-289 FROM N.A.
 RX STRAIN-K12;
 RX MEDLINE=87115164; PubMed=3027506;
 RA Irlino N., Nakayama K., Nakayama H.;
 RT "The reco gene of Escherichia coli K12: primary structure and
 RT evidence for SOS regulation.";
 RL Mol. Gen. Genet. 205:298-304(1986).
 RN 18
 RP MUTAGENESIS OF SER-172.
 RX MEDLINE=9431966; PubMed=8300539;
 RA Brok R.C.P.M., Brinkman E., van Boxtel R., Bekkers A.C.A.P.,
 RA Verheij H.M., Tomassen J.;
 RT "Molecular characterization of enterobacterial plda genes encoding
 RT outer membrane phospholipase A.";
 RL J. Bacteriol. 176:861-870(1994).
 RN 19
 RP ACTIVE SITE SER-164, AND PARTIAL SEQUENCE.
 RX MEDLINE=91249806; PubMed=2040286;
 RA Horrevorts A.J.G., Verheij H.M., de Haas G.H.;
 RT "Inactivation of Escherichia coli outer-membrane phospholipase A by
 RT the affinity label hexadecanesulfonyl fluoride. Evidence for an
 RT active-site serine.";
 RL Eur. J. Biochem. 198:247-253(1991).
 CC -1- FUNCTION: HYDROLYSIS OF PHOSPHATIDYLCHOLINE WITH PHOSPHOLIPASE
 CC A2 (EC 3.1.1.4) AND PHOSPHOLIPASE A1 (EC 3.1.1.32) ACTIVITIES.
 CC REQUIRED FOR EFFICIENT SECRETION OF BACTERIOCINS, SEEMS TO BE
 CC DORMANT IN NORMAL GROWING CELLS.
 CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 1-ACYLGLYCERO-
 CC PHOSPHOCHOLINE + A FATTY ACID ANION.
 CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 2-ACYLGLYCERO-
 CC PHOSPHOCHOLINE + A FATTY ACID ANION.
 CC -1- COFACTOR: REDUCES CALCIUM IONS FOR ACTIVITY.
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE; ONE OF THE VERY FEW ENZYMES
 CC LOCATED THERE.
 CC -1- INDUCTION: BY MEMBRANE DAMAGING, FOR EXAMPLE, BY PHAGE-INDUCED
 CC LYSIS OR TEMPERATURE SHOCK.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL, outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X02143; CAA26081.1; -;
DR EMBL: M87049; AAA67617.1; -;
DR EMBL: AE000458; AAC76824.1; -;
DR EMBL: AE005613; AAG59017.1; -;
DR EMBL: AP002567; BAB38174.1; -;
DR EMBL: M30198; AAA24516.1; -;
DR PIR: A00771; PSECA.
DR PIR: A22133; PSECA1.
DR PIR: S30711; S30711.
DR EcoGene: EG10738; PldA.
DR InterPro: IPR003187; PLA1.
DR Pfam: PF02253; PLA1; 1.
KW Hydrolyase; Lipid degradation; Outer membrane; Signal; Calcium;
KW Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 289 PHOSPHOLIPASE A1.
FT ACT_SITE 164 164
FT MUTAGEN 172 172 S->F: INACTIVE PROTEIN.
FT CONFLICT 14 15 LP -> FA (IN REF. 2).
FT CONFLICT 30 33 DAPA -> MTRQ (IN REF. 6).
SQ SEQUENCE 289 AA: 33163 MW: A68BAD32AA60F218 CRC64;

Query Match 93.38; Score 28; DB 1; Length 289;
Best Local Similarity 80.0%; Pred. No. 97;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FPXWR 5
11 11
DB 95 FPLWR 99

Search completed: February 27, 2002, 11:42:48
Job time: 347 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:50:12 ; Search time 281.76 Seconds
(without alignments)
2.596 Million cell updates/sec

Title: US-09-446-109a-13

Perfect score: 30

Sequence: 1 FPXMR 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_PROTIST:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	96.7	101	5	Q9V671 drosophila
2	29	96.7	113	10	Q39293 brassica na
3	29	96.7	157	5	Q9XRV5 caenorhabd
4	29	96.7	158	4	Q9NMW2 homo sapien
5	29	96.7	158	4	Q9NMW2 homo sapien
6	29	96.7	229	8	Q37633 rhagoletis
7	29	96.7	245	2	Q9PDE1
8	29	96.7	273	10	Q9SU09
9	29	96.7	292	2	Q9H783
10	29	96.7	321	11	Q9D6L7
11	29	96.7	332	2	Q9APJ5
12	29	96.7	333	2	Q9L3G9
13	29	96.7	334	10	Q9FTM4
14	29	96.7	349	4	Q15429 homo sapien
15	29	96.7	417	10	Q9SAA1
16	29	96.7	466	2	Q9FCK4
17	29	96.7	479	2	Q86742 zymomonas m
18	29	96.7	497	2	Q91702 streptomyces
19	29	96.7	501	5	Q16923 caenorhabd

20	29	96.7	502	5	Q9GUIS	Q9gu15 caenorhabd
21	29	96.7	555	1	Q29748	Q29748 archaeglob
22	29	96.7	581	4	Q9BRQ1	Q9brq1 homo sapien
23	29	96.7	612	4	Q9H8Y2	Q9h8y2 homo sapien
24	29	96.7	621	4	Q9H9Y1	Q9h9y1 homo sapien
25	29	96.7	625	2	Q66676	Q66676 aquifex aco
26	29	96.7	634	4	Q9MM19	Q9mm19 homo sapien
27	29	96.7	634	4	Q9NMX5	Q9nmx5 homo sapien
28	29	96.7	642	2	Q25396	Q25396 helicobacte
29	29	96.7	642	2	Q9ZLF3	Q9zlf3 helicobacte
30	29	96.7	650	5	Q17866	Q17866 caenorhabd
31	29	96.7	735	2	Q9HU63	Q9hu63 pseudomonas
32	29	96.7	778	5	Q9U9K6	Q9u9k6 caenorhabd
33	29	96.7	811	4	Q9BMX2	Q9bm2 homo sapien
34	29	96.7	1071	11	Q9D2K4	Q9d2k4 mus musculu
35	29	96.7	1108	13	Q9PMD0	Q9pmd0 tetraodon f
36	29	96.7	1117	5	Q9U9K7	Q9u9k7 caenorhabd
37	29	96.7	1451	2	Q45336	Q45336 bordetella
38	29	96.7	1451	2	Q45044	Q45044 bordetella
39	29	96.7	1464	2	Q9S5D5	Q9s5d5 bordetella
40	28	93.3	68	2	P73858	P73858 synecocyst
41	28	93.3	84	2	P74471	P74471 synecocyst
42	28	93.3	84	2	O55568	O55568 synecocyst
43	28	93.3	93	2	P73028	P73028 synecocyst
44	28	93.3	103	10	Q9M045	Q9m045 arabidopsis
45	28	93.3	110	6	Q9GLJ9	Q9glj9 canis famli

ALIGNMENTS

RESULT 1
ID Q9V671 PRELIMINARY: PRT: 101 AA.
AC Q9V671
DT 01-MAY-2000 (TREMBLrel. 13, Created)
UT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
ET 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CG13174 PROTEIN.
GN CG13174.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:
OC Pterygota: Neoptera: Endopterygota: Diptera: Brachycera: Muscomorpha:
OC Ephydroidea: Drosophilidae: Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKLEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Moriman J.R., Yandell M.D., Zhang O., Chen L.X.,
RA Brandon R., Rogers Y.H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Ayale C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aydayani A., An H.-J., Andrews-Planckoch C., Baldwin D.,
RA Baillet R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu F., Center A., Chandra I.,
RA Cherry J.M., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson B., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Correll J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.K., Houck J.,
RA Hostlin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagham C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Palazotto K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissendach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003823; AAF58560.1; -;
 DR FlyBase: FBgn0033694; CG13174.
 SQ SEQUENCE 101 AA: 11820 MW: 1497830707223948 CRC64;

Query Match 96.7%; Score 29; DB 5; Length 101;
 Best Local Similarity 80.0%; Pred. No. 1.2e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FPXMR 5
 DB 75 FPMWR 79

RESULT 2
 Q39293 PRELIMINARY: PRT; 113 AA.
 AC Q39293;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE ELONGATION FACTOR EF-1A (FRAGMENT).
 OS Brassica napus (Rape).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Brassica.
 OX NCBL_TaxID=3708;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SAMOURL; TISSUE=ETIOLATED SEEDLINGS, (5-DAYS-OLD);
 RA Saenz-Vasquez J.;
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO GTP-BINDING ELONGATION FACTOR FAMILY.
 DR EMBL: U21744; AAA6366.1; -;
 DR Mendel: 15767; Brana:1139;15767.
 DR InterPro: IPR00795; GTP_EFTU.
 DR Pfam: PF00009; GTP_EFTU; 1.
 DR Elongation factor: GTP-binding; Protein biosynthesis.
 KW NON TER
 FT 1
 SQ SEQUENCE 113 AA: 12560 MW: 1497830707223948 CRC64;

Query Match 96.7%; Score 29; DB 10; Length 113;
 Best Local Similarity 80.0%; Pred. No. 1.4e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FPXMR 5
 DB 103 FPMWR 107

RESULT 3
 Q9XTY5 PRELIMINARY: PRT; 157 AA.
 AC Q9XTY5;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE T03D8.2 PROTEIN.

GN T03D8.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBL_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Mortimore B.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed-7906398;
 RA Willson R., Almscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rikken L., Roopra A., Saunders D., Showkhen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterson R.,
 RA Watson A., Weinstein L., Wilkinson-Sproat J., Wohlman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*.";
 RL Nature 368:32-38(1994).
 DR EMBL: Z92838; CAB07406.1; -;
 DR InterPro: IPR00230; Ribosomal_S12.
 DR Pfam: PF00164; Ribosomal_S12; 1.
 DR PRINTS: PR01034; RIBOSOMALS12.
 DR PRODOM: PD000576; Ribosomal_S12; 1.
 SQ SEQUENCE 157 AA: 17370 MW: 6064837282E826C CRC64;

Query Match 96.7%; Score 29; DB 5; Length 157;
 Best Local Similarity 80.0%; Pred. No. 1.9e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FPXMR 5
 DB 40 FPMWR 44

RESULT 4
 Q9NMW3 PRELIMINARY: PRT; 158 AA.
 AC Q9NMW3;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE MEMBRANE COFACTOR PROTEIN CD46 VARIANT (FRAGMENT).
 OS MCP.
 GN Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBL_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20227770; PubMed-10751138;
 RA Kusutara K., Sasaki Y., Nakao F., Ihara K., Hattori H., Yamashita S.,
 RA Nihel K., Koide N., Aiba H., Takeshita K., Hara T.;
 RT "Analysis of measles virus binding sites of the CD46 gene in patients
 with subacute sclerosing panencephalitis.";
 RL J. Infect. Dis. 181:1447-1449(2000).
 DR EMBL: AF209713; AAF73845.1; -;
 DR InterPro: IPR000436; Sush1-SCR_CCP.
 DR Pfam: PF00084; sush1; 2.
 DR SMART: SM00032; CCP; 2.
 FT VARIANT 59 59 Q -> R.
 FT NON TER 158 158
 SQ SEQUENCE 158 AA: 18098 MW: 4E307EEC5C454F5 CRC64;

Query Match 96.7%; Score 29; DB 4; Length 158;

Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
1111
DB 11 FPSMR 15

RESULT 5

Q9NNW2 PRELIMINARY; PRT: 158 AA.

AC Q9NNW2;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE MEMBRANE COFACTOR PROTEIN CD46 VARIANT (FRAGMENT).
MCP.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN 11

RP SEQUENCE FROM N.A.

RC MEDLINE-2027770; PubMed-10751138;

RA Kusunoha K., Sasaki Y., Nakao F., Ihara K., Hattori H., Yamashita S.,
Nihei K., Koide N., Alba H., Takeshita K., Hara T.;

RT "Analysis of measles virus binding sites of the CD46 gene in patients
with subacute sclerosing panencephalitis.";

RL J. Infect. Dis. 181:1447-1449(2000).

DR EMBL; AF209714; AAF73846.1; -;

DR InterPro: IPR000436; Sushl_SCR_CCP.

DR Pfam: PF00084; sushl. 2.

DR SMART: SM00032; CCP; 2.

FT NON_TER 158

FT SEQUENCE 158 AA; 18126 MW; E2CF38EC3A16A9DA CRC64;

Query Match 96.7%; Score 29; DB 4; Length 158;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
1111
DB 11 FPSMR 15

RESULT 6
Q37633 PRELIMINARY; PRT: 229 AA.

AC Q37633;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE CYTOCHROME C OXIDASE POLYPEPTIDE II (BC 1.9.3.1) (FRAGMENT).

GN COXI.

OS Rhagoletis boycel.

OC Mitochondrion.

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Tephritidae; Tephritidae; Rhagoletis.

OX NCBI_TaxID=43419;

RN 11

RP SEQUENCE FROM N.A.

RC MEDLINE-97159559; PubMed-9007018;

RA Smith J.J., Bush G.L.;

RT "Phylogeny of the genus Rhagoletis (Diptera: Tephritidae) inferred
from DNA sequences of mitochondrial cytochrome oxidase II.";

RL Mol. Phylogenet. Evol. 7:33-43(1997).

DE - FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 2

CC FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2

CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER

CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1 (BY

CC SIMILARITY).

CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) - 2 H(2)O + 4

CC FERRICYTOCHROME C.

CC -1- COFACTOR: COPPER A AND HEME GROUP.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL

CC INNER MEMBRANE (BY SIMILARITY).

CC -1- SIMILARITY: TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.

DR EMBL; U53254; AAB50355.1; -;

DR HSSP: P08306; IARI.

DR InterPro: IPR001505; COX2.

DR InterPro: IPR002429; Cyt_c-ox_2.

DR Pfam: PF00116; COX2; 1.

DR PRINTS: PR01166; CYCOXIDASE11.

DR ProDom: PD000131; COX2; 1.

DR PROSITE: PS00078; COX2; 1.

KW Copper; Inner membrane; Mitochondrion; Oxidoreductase;

KW Respiratory chain; Transmembrane.

FT METAL 159 159 COPPER A (PROBABLE).

FT METAL 163 163 COPPER A (PROBABLE).

FT METAL 167 167 COPPER A (PROBABLE).

FT METAL 207 207 COPPER A (PROBABLE).

FT NON_TER 229

FT SEQUENCE 229 AA; 26389 MW; 912D1B2736330698 CRC64;

Query Match 96.7%; Score 29; DB 8; Length 229;
Best Local Similarity 80.0%; Pred. No. 2.7e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
1111
DB 78 FPSMR 82

RESULT 7
Q9PDE1 PRELIMINARY; PRT: 245 AA.

AC Q9PDE1;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)

DE HYPOTHETICAL PROTEIN XF1438.

GN XF1438.

OS Xylella fastidiosa.

OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;

OC Xylella.

OX NCBI_TaxID=2371;

RN 11

RP SEQUENCE FROM N.A.

RC STRAIN=9A5C.

RC MEDLINE-20365717; PubMed-10910347;

RA Simpson A.J.G., Reinuch F.C., Arruda P., Abreu F.A., Acencio M.,
Alvares A.R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,

Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Britones M.R.S.,

Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carreir H.,

Colancho N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,

Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,

Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,

Fraga J.S., Franca S.C., Franco M.C., Froime M., Furlan L.R.,

Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,

Ho P.L., Hohnselt J.D., Jungueira M.L., Kemper E.L., Kitajima J.P.,

Kriegler J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,

Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,

Machado M.A., Madeira A.M.B.N., Martins E.M.F., Matsunuma A.T.,

Marques M.V., Martins A.A.L., Martins E.M.F., Matsunuma A.T.,

Menck C.F.M., Miracca E.C., Miyaki C.V., Monteiro-Vitorello G.B.,

Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,

Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,

de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,

Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,

Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,

de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sasaki H.E.,

da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,

RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Trufi D., Tsai S.M., Tsubako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovsky-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
 .RT "The genome sequence of the plant pathogen *Xylella fastidiosa*.";
 .RL Nature 406:151-159(2000).
 DR EMBL: AE003974; AAF84247.1; -
 KW Hypothetical protein; Complete proteome.
 SO SEQUENCE 245 AA; 27193 MW; 16C8ADB9FC490455 CRC64;

Query Match: 96.7%; Score 29; DB 2; Length 245;
 Best Local Similarity 80.0%; Pred. No. 2.8e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 DB 197 FPSMR 201

RESULT 8
 ID 09SU9 PRELIMINARY: PRT: 273 AA.
 AC 09SU9:
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE HYPOTHEICAL 31.2 KDA PROTEIN.
 GN F7K2.180 OR ATG4G23600.
 GN Arabidopsis thaliana (Mouse-ear cress).
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 RX NCBI_TaxID=3702;
 RN 11
 RP SEQUENCE FROM N.A.
 RA Bevan M., Medler H., Wambutt R., Bancroft I., Mewes H.W.,
 RA Mayer K.F.X., Lemcke K., Schueller C.;
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RN 12
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 RN 13
 RP SEQUENCE FROM N.A.
 RA Medler H., Wambutt R., Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN 14
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL033545; CNA22165.1; -
 DR EMBL: AL161557; CAB79215.1; -
 KM Hypothetical protein.
 SO SEQUENCE 273 AA; 31248 MW; 76644F68CE8EBB0 CRC64;

Query Match: 96.7%; Score 29; DB 10; Length 273;
 Best Local Similarity 80.0%; Pred. No. 3.1e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 DB 79 FPSMR 83

RESULT 9
 ID 09HT83 PRELIMINARY: PRT: 292 AA.
 AC 09HT83:
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE HYPOTHETICAL PROTEIN PA5488.
 GN PA5488.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 RX NCBI_TaxID=287;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Huffnagle W.O., Kovallik D.J., Lagrou M.,
 RA Gardner R.L., Collier L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Polger K.R., Kas A., Larling K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.F.W., Lory S., Olson M.V.;
 .RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an
 opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 DR EMBL: AE004961; AAG08873.1; -
 DR InterPro: IPR000504; RRM.
 DR PROSITE: PS00030; RRM_RNP_1; UNKNOWN_1.
 KW Hypothetical protein; Complete proteome.
 SO SEQUENCE 292 AA; 32949 MW; 14610337EBCCFD4 CRC64;

Query Match: 96.7%; Score 29; DB 2; Length 292;
 Best Local Similarity 80.0%; Pred. No. 3.3e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 DB 231 FPSMR 235

RESULT 10
 ID 09D6L7 PRELIMINARY: PRT: 321 AA.
 AC 09D6L7:
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE 231007601ARIK PROTEIN.
 GN 231007601ARIK.
 GN Mus musculus (Mouse).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RX NCBI_TaxID=10090;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TONGUE;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Atkawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Koehli H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner I., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming I.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,
 RA Hayashizaki Y.;
 .RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL: AK010208; BAB26769.1; -

```
DR MCD:1919189, 2310076014RIK.
DR InterPro: IPR000847; HTH_LysR.
DR PROSITE: PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.
SQ SEQUENCE 321 AA: 35038 MW: 6E27EBC746552DFB CRC64:

Query Match
Best Local Similarity 96.7%; Score 29; DB 11; Length 321;
Matches 4; Conservative 0; Pred. No. 3.7e+02; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
DB 44 FPMWR 48

RESULT 11
O9APJ5 PRELIMINARY: PRT; 332 AA.
AC O9APJ5;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ENDO-1,4-BETA-D-GLUCANASE PRECURSOR.
CN CEL8Y.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=556;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=py35;
RA "Cho S.-J., Park S.R., Yun H.D.;
RT "Cloning and sequencing of a cel8Y gene of Pectobacterium chrysanthemi
RT py35."
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF282321; AAG49556.1; -.
KW SIGNAL.
FT SIGNAL.
SQ SEQUENCE 332 AA: 37627 MW: A1B5D7B0CBB20EE6 CRC64:

Query Match
Best Local Similarity 96.7%; Score 29; DB 2; Length 332;
Matches 4; Conservative 0; Pred. No. 3.8e+02; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
DB 178 FPMWR 182

RESULT 12
O9LJG9 PRELIMINARY: PRT; 333 AA.
AC O9LJG9;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4).
CN CELA.
OS Erwinia rhapsodict.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Erwinia.
OX NCBI_TaxID=55212;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCPPB2989;
RA Saarialhti H.T.;
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NCPPB2989;
RA Rieckl R.;
RT "Members of the amylovora group of Erwinia are cellulytic and
```

```
RT possess genes homologous to the type II secretion pathway."
RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AJ276358; CAB89803.1; -.
DR InterPro: IPR002037; Glyco_hydro_8.
DR Pfam: PF01270; Glyco_hydro_8; 1.
DR PRINTS: PR00735; GLYHIDRASE8.
KW Signal; Hydrolase; Glycosidase.
FT SIGNAL.
FT CHAIN 1 23
SQ SEQUENCE 333 AA: 37783 MW: D75CFE212302673A CRC64:

Query Match
Best Local Similarity 96.7%; Score 29; DB 2; Length 333;
Matches 4; Conservative 0; Pred. No. 3.8e+02; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
DB 178 FPMWR 182

RESULT 13
O9FTM4 PRELIMINARY: PRT; 334 AA.
AC O9FTM4;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE P0005A05.25 PROTEIN.
CN Oryza sativa (Rice).
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0005A05."
RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AP002863; BAB16921.1; -.
DR InterPro: IPR000520; Exonuclease.
DR InterPro: IPR000822; Znf-C2H2.
DR Pfam: PF00929; Exonuclease; 1.
DR Pfam: PF00096; Zf-C2H2; 1.
DR SMART: SM00479; EXO11; 1.
DR SMART: SM00355; ZNF_C2H2; 2.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 1.
KW DNA-binding; Zinc-finger.
SQ SEQUENCE 334 AA: 37172 MW: A34531D625704A88 CRC64:

Query Match
Best Local Similarity 96.7%; Score 29; DB 10; Length 334;
Matches 4; Conservative 0; Pred. No. 3.8e+02; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
DB 301 FPMWR 305

RESULT 14
O15429 PRELIMINARY: PRT; 349 AA.
AC O15429;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CD46.
OS Homo sapiens (Human).
```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=TESTIS;
 RA Hara T., Suzuki Y., Nagasawa S., Seya T.;
 RT "CDNA cloning and characterization of human sperm CD46."
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: D84105; BAA12224.1; -
 DR HSSP: P10998; 1VD.
 DR Interpro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF00084; sushi; 4.
 DR SMART: SM0032; CCP; 4.
 SQ SEQUENCE 349 AA: 39325 MW: 8EFCEDA30D3C818E CRC64;

Query Match 96.7%; Score 29; DB 4; Length 349;
 Best Local Similarity 80.0%; Pred. No. 4e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FPXWR 5
 1 1 1
 Db 11 FPXWR 15

RESULT 15
 O9SA1
 ID O9SA1 PRELIMINARY; PRT; 417 AA.
 AC O9SA1:
 DT 01-MAY-2000 (TREMREL. 13, Created)
 DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
 DT 01-MAY-2000 (TREMREL. 13, Last annotation update)
 DE F25C20.9.
 GN F25C20.9.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
 RA Li J., Lee J.M., Kremenetskaia I., Lufos J., Ngan I., Liu A.,
 RA Gonzalez A., Altali H., Araujo R., Chao Q., Conn L., Conway A.B.,
 RA Dunn P., Hansen N., Huizar L., Kim C., Palm C., Rowley D., Shinn P.,
 RA Walker M., Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
 RT "Arabidopsis thaliana chromosome 1 BAC F25C20 sequence."
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Theologis A.;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Theologis A.;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC007296; AAD30247.1; -
 SQ SEQUENCE 417 AA: 45985 MW: B86EBF06AC19A63D CRC64;

Query Match 96.7%; Score 29; DB 10; Length 417;
 Best Local Similarity 80.0%; Pred. No. 4.7e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FPXWR 5
 1 1 1
 Db 377 FPXWR 381

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:36:06 : Search time 132.19 seconds
(without alignments)
0.851 Million cell updates/sec

Title: US-09-446-109a-13
Perfect score: 30
Sequence: 1 FPXMR 5

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: Issued_Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCrUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	96.7	38	1	US-08-176-500-118 Sequence 118, App
2	29	96.7	38	1	US-08-471-052A-118 Sequence 118, App
3	29	96.7	38	1	US-08-189-331-118 Sequence 118, App
4	29	96.7	38	2	US-08-471-939-118 Sequence 118, App
5	29	96.7	38	2	US-08-471-800-118 Sequence 118, App
6	29	96.7	38	2	US-08-471-068-118 Sequence 118, App
7	29	96.7	241	4	US-08-823-120-1 Sequence 1, Appl
8	29	96.7	324	2	US-08-528-057-46 Sequence 46, Appl
9	29	96.7	370	2	US-08-528-057-42 Sequence 42, Appl
10	29	96.7	372	1	US-08-202-056-5 Sequence 5, Appl
11	29	96.7	372	1	US-08-076-093A-6 Sequence 6, Appl
12	29	96.7	372	1	US-08-701-265-6 Sequence 6, Appl
13	29	96.7	372	2	US-08-284-586-6 Sequence 6, Appl
14	29	96.7	372	2	US-08-805-478-6 Sequence 6, Appl
15	29	96.7	372	2	US-08-802-627A-6 Sequence 6, Appl
16	29	96.7	372	2	US-08-801-238-6 Sequence 6, Appl
17	29	96.7	372	2	US-08-801-328-6 Sequence 6, Appl
18	29	96.7	372	3	US-08-104-298-6 Sequence 6, Appl
19	29	96.7	372	3	US-08-982-493-8 Sequence 8, Appl
20	29	96.7	373	2	US-08-528-057-44 Sequence 44, Appl
21	29	96.7	377	2	US-08-528-057-2 Sequence 2, Appl
22	29	96.7	384	4	US-08-139-195-2 Sequence 2, Appl
23	29	96.7	384	6	5514787-2 Patent No. 5514787
24	28	93.3	57	1	US-08-370-225-29 Sequence 29, Appl
25	28	93.3	57	1	US-08-370-225-30 Sequence 30, Appl
26	28	93.3	57	1	US-08-461-859-29 Sequence 29, Appl
27	28	93.3	57	1	US-08-461-859-30 Sequence 30, Appl

28	28	93.3	57	5	PCT-US93-10069-29	Sequence 29, Appl
29	28	93.3	57	5	PCT-US93-10069-30	Sequence 30, Appl
30	28	93.3	298	2	US-08-061-636-3	Sequence 3, Appl
31	28	93.3	298	2	US-08-874-347-19	Sequence 19, Appl
32	28	93.3	298	3	US-09-093-522-19	Sequence 19, Appl
33	28	93.3	298	3	PCT-US94-05268-3	Sequence 3, Appl
34	28	93.3	373	2	US-08-846-762-13	Sequence 13, Appl
35	28	93.3	374	3	US-08-982-493-6	Sequence 6, Appl
36	28	93.3	437	2	US-09-031-059-1	Sequence 1, Appl
37	28	93.3	437	2	US-09-031-059-3	Sequence 3, Appl
38	28	93.3	760	1	US-08-195-152-2	Sequence 2, Appl
39	26	86.7	10	4	US-09-461-697-409	Sequence 409, App
40	26	86.7	12	4	US-09-461-697-407	Sequence 407, App
41	26	86.7	113	1	US-07-668-648-10	Sequence 10, Appl
42	26	86.7	113	2	US-08-429-998-10	Sequence 10, Appl
43	26	86.7	113	2	US-08-431-333-10	Sequence 10, Appl
44	26	86.7	113	5	PCT-US91-02321-10	Sequence 10, Appl
45	26	86.7	862	1	US-08-325-267A-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-176-500-118
Sequence 118, Application US/08176500
Patent No. 5498538
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
APPLICANT: Fowlkes, D. M.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 141
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/176,500
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/013,416
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101:143
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-176-500-118

Query Match 96.74: Score 29: DB 1: Length 38:
Best Local Similarity 80.08: Pred. No. 19:
Matches 4: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

OY 1 FPXMR 5
11 11
Db 10 FPSMR 14

RESULT 2
US-08-471-052A-118

Sequence 118, Application US/08471052A

Patent No. 5625033

GENERAL INFORMATION:

APPLICANT: Kay, B. K.

APPLICANT: Fowlkes, D. M.

TITLE OF INVENTION: Totally Synthetic Affinity Reagents

NUMBER OF SEQUENCES: 166

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/471,052A

FILING DATE: 06-JUNE-1995

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 1101-179

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212 790-9090

TELEFAX: 212 869-8864/9741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 118:

SEQUENCE CHARACTERISTICS:

LENGTH: 38 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-471-052A-118

Query Match 96.7%; Score 29; DB 1; Length 38;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
11 11
Db 10 FPSMR 14

RESULT 3
US-08-189-331-118

Sequence 118, Application US/08189331

Patent No. 5747334

GENERAL INFORMATION:

APPLICANT: Kay, B. K.

APPLICANT: Fowlkes, D. M.

TITLE OF INVENTION: Totally Synthetic Affinity Reagents

NUMBER OF SEQUENCES: 166

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/189,331

FILING DATE: Concurrently herewith

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 1101-155

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212 790-9090

TELEFAX: 212 869-8864/9741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 118:

SEQUENCE CHARACTERISTICS:

LENGTH: 38 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-189-331-118

Query Match 96.7%; Score 29; DB 1; Length 38;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
11 11
Db 10 FPSMR 14

RESULT 4
US-08-471-939-118

Sequence 118, Application US/08471939

Patent No. 5844076

GENERAL INFORMATION:

APPLICANT: Kay, B. K.

APPLICANT: Fowlkes, D. M.

TITLE OF INVENTION: Totally Synthetic Affinity Reagents

NUMBER OF SEQUENCES: 141

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/471,939

FILING DATE: 06-JUN-1995

CLASSIFICATION: 530

PRIOR APPLICATION NUMBER:

APPLICATION NUMBER: US 08/013,416

FILING DATE: 01-FEB-1993

ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 1101-143

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212 790-9090

TELEFAX: 212 869-8864/9741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 118:

SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-471-939-118

Query Match 96.7%; Score 29; DB 2; Length 38;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
DB 10 FPSWR 14

RESULT 5
US-08-471-800-118
Sequence 118, Application US/08471800
Patent No. 5852167
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
APPLICANT: Fowlkes, D. M.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 141
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennile & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471.800
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/013,416
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-143
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-471-800-118

Query Match 96.7%; Score 29; DB 2; Length 38;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
DB 10 FPSWR 14

RESULT 6
US-08-471-068-118
Sequence 118, Application US/08471068
Patent No. 5948635
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
APPLICANT: Fowlkes, D. M.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 186
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennile & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,068
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/189,331
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-155
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-471-068-118

Query Match 96.7%; Score 29; DB 2; Length 38;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
DB 10 FPSWR 14

RESULT 7
US-08-823-120-1
Sequence 1, Application US/08823120
Patent No. 6149919
GENERAL INFORMATION:
APPLICANT: Domenighini, Mario
APPLICANT: Rappuoli, Rino
APPLICANT: Pizza, Mariagrazia
TITLE OF INVENTION: Immunogenic Detoxified Mutants of
Cholera Toxin and of the Toxin Lt, Their Preparation and
TITLE OF INVENTION: Their Use for the Preparation of Vaccines
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: California
COUNTRY: USA

ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/823,120
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/256,003
FILING DATE: 11-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0315,001
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 241 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-823-120-1

Query Match 96.7%; Score 29; DB 4; Length 241;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FPXMR 5
DB 169 FPMWR 173

RESULT 8
US-08-528-057-46
Sequence 46, Application US/08528057
Patent No. 5846715
GENERAL INFORMATION:
APPLICANT: PURCELL, Damian F. J.
APPLICANT: RUSSELL, Sarah M.
APPLICANT: MCKENZIE, Ian F. C.
TITLE OF INVENTION: CD46 VARIANTS
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/528,057
FILING DATE: CONCURRENTLY HERewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/961,686
FILING DATE: 11-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU91/00199
FILING DATE: 10-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PK0133/90
FILING DATE: 11-MAY-1990

ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 17227/112 DACO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 324 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-528-057-46

Query Match 96.7%; Score 29; DB 2; Length 324;
Best Local Similarity 80.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FPXMR 5
DB 11 FPMWR 15

RESULT 9
US-08-528-057-42
Sequence 42, Application US/08528057
Patent No. 5846715
GENERAL INFORMATION:
APPLICANT: PURCELL, Damian F. J.
APPLICANT: RUSSELL, Sarah M.
APPLICANT: MCKENZIE, Ian F. C.
TITLE OF INVENTION: CD46 VARIANTS
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/528,057
FILING DATE: CONCURRENTLY HERewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/961,686
FILING DATE: 11-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU91/00199
FILING DATE: 10-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PK0133/90
FILING DATE: 11-MAY-1990
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 17227/112 DACO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 370 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-528-057-42

Query Match 96.7%; Score 29; DB 1; Length 370;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
11 11
DB 11 FPSMR 15

RESULT 10
US-08-202-056-5
Sequence 5, Application US/08202056
Patent No. 5440021
GENERAL INFORMATION:
APPLICANT: Chuntharapai, Anan
APPLICANT: Hebert, Caroline
APPLICANT: Kim, Kyung Jin
APPLICANT: Lee, James
TITLE OF INVENTION: Antibodies to Human IL-8 Type B Receptor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202,056
FILING DATE: 25-FEB-1994
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/677211
FILING DATE: 29-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B.
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: 706P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-202-056-5

Query Match 96.7%; Score 29; DB 1; Length 372;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
11 11
DB 352 FPSMR 356

RESULT 11
US-08-076-093A-6
Sequence 6, Application US/08076093A
Patent No. 5543503
GENERAL INFORMATION:

APPLICANT: Chuntharapai, Anan
APPLICANT: Lee, James
APPLICANT: Hebert, Caroline
APPLICANT: Jin Kim, K.
TITLE OF INVENTION: Antibodies to Human PFAA Receptors
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/076,093A
FILING DATE: 11-Jun-1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/677211
FILING DATE: 29-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B.
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: 706P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: linear
US-08-076-093A-6

Query Match 96.7%; Score 29; DB 1; Length 372;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
11 11
DB 352 FPSMR 356

RESULT 12
US-08-701-265-6
Sequence 6, Application US/08701265
Patent No. 5776457
GENERAL INFORMATION:
APPLICANT: Chuntharapai, Anan
APPLICANT: Lee, James
APPLICANT: Hebert, Caroline
APPLICANT: Jin Kim, K.
TITLE OF INVENTION: Antibodies to Human PFAA Receptors
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/701,265
FILING DATE: 22-AUG-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/076093
FILING DATE: 11-JUN-1993
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/677211
FILING DATE: 29-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: 706P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-701-265-6

Query Match 96.7%; Score 29; DB 1; Length 372;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PPXMR 5
1111
DB 352 PPSWR 356

RESULT 13
US-08-284-586-6
Sequence 6, Application US/08284586
GENERAL INFORMATION:
APPLICANT: Chuntharapal, Anan
APPLICANT: Lee, James
APPLICANT: Hebert, Caroline
APPLICANT: Jin Kim, K.
TITLE OF INVENTION: Antibodies to Human PFAA Receptors
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/284,586
FILING DATE:
NAME: Love, Richard B
REGISTRATION NUMBER: 424
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/677211
FILING DATE: 29-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: 706P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-284-586-6

Query Match 96.7%; Score 29; DB 2; Length 372;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PPXMR 5
1111
DB 352 PPSWR 356

RESULT 14
US-08-805-478-6
Sequence 6, Application US/08805478
GENERAL INFORMATION:
APPLICANT: Chuntharapal, Anan
APPLICANT: Lee, James
APPLICANT: Hebert, Caroline
APPLICANT: Jin Kim, K.
TITLE OF INVENTION: ANTIBODIES TO PFAA RECEPTOR
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/805,478
FILING DATE: 25-Feb-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/284586
FILING DATE: 10-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/076093
FILING DATE: 11-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: P0706P2P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:

LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-805-478-6

Search completed: February 27, 2002, 11:36:07
Job time: 146 sec

Query Match 96.7%; Score 29; DB 2; Length 372;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
DB 352 FPSMR 356

RESULT 15
US-08-802-627A-6
Sequence 6, Application US/08802627A
Patent No. 5892017
GENERAL INFORMATION:
APPLICANT: Lee, James
APPLICANT: Wood, William I.
TITLE OF INVENTION: NUCLEIC ACID ENCODING PFAA RECEPTOR
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Minipalin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/802,627A
FILING DATE: 19-Feb-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/284586
FILING DATE: 10-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/076093
FILING DATE: 11-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B.
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: P0706P2P1D2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-802-627A-6

Query Match 96.7%; Score 29; DB 2; Length 372;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
DB 352 FPSMR 356

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:41:18 : Search time 303.5 Seconds
(without alignments)
1.220 Million cell updates/sec

Title: US-09-446-109A-14

Perfect score: 30

Sequence: 1 FPKWR 5

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_1101.*
1: /SID8/gcgdata/geneseq/geneseqp/AA1980.DAT.*
2: /SID8/gcgdata/geneseq/geneseqp/AA1981.DAT.*
3: /SID8/gcgdata/geneseq/geneseqp/AA1982.DAT.*
4: /SID8/gcgdata/geneseq/geneseqp/AA1983.DAT.*
5: /SID8/gcgdata/geneseq/geneseqp/AA1984.DAT.*
6: /SID8/gcgdata/geneseq/geneseqp/AA1985.DAT.*
7: /SID8/gcgdata/geneseq/geneseqp/AA1986.DAT.*
8: /SID8/gcgdata/geneseq/geneseqp/AA1987.DAT.*
9: /SID8/gcgdata/geneseq/geneseqp/AA1988.DAT.*
10: /SID8/gcgdata/geneseq/geneseqp/AA1989.DAT.*
11: /SID8/gcgdata/geneseq/geneseqp/AA1990.DAT.*
12: /SID8/gcgdata/geneseq/geneseqp/AA1991.DAT.*
13: /SID8/gcgdata/geneseq/geneseqp/AA1992.DAT.*
14: /SID8/gcgdata/geneseq/geneseqp/AA1993.DAT.*
15: /SID8/gcgdata/geneseq/geneseqp/AA1994.DAT.*
16: /SID8/gcgdata/geneseq/geneseqp/AA1995.DAT.*
17: /SID8/gcgdata/geneseq/geneseqp/AA1996.DAT.*
18: /SID8/gcgdata/geneseq/geneseqp/AA1997.DAT.*
19: /SID8/gcgdata/geneseq/geneseqp/AA1998.DAT.*
20: /SID8/gcgdata/geneseq/geneseqp/AA1999.DAT.*
21: /SID8/gcgdata/geneseq/geneseqp/AA2000.DAT.*
22: /SID8/gcgdata/geneseq/geneseqp/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	96.7	10	AAV08336	O. hannah venom al
2	29	96.7	11	AAV08334	O. hannah venom al
3	29	96.7	38	AAV08491	Binding domain of
4	29	96.7	94	AAV06599	Rat sodium channel
5	29	96.7	94	AAV020130	Rat truncated sodi
6	29	96.7	135	AAV04918	Mycobacterium spec
7	29	96.7	177	AAV058892	Breast and ovarian
8	29	96.7	299	AAV052488	Helicobacter pylori
9	29	96.7	314	AAV03943	CD46 construct del
10	29	96.7	324	AAV15233	CD46 from clone pm
11	29	96.7	359	AAV0396	Cellulase. Cellul

12	29	96.7	372	13	AAV27793	New platelet facto
13	29	96.7	372	16	AAV92239	Chemokine superfam
14	29	96.7	372	16	AAV68813	Human lymphocyte P
15	29	96.7	372	20	AAV06644	Human Burkitt's ly
16	29	96.7	372	21	AAV90627	Human G protein-co
17	29	96.7	372	21	AAV90661	Human mutant G pro
18	29	96.7	372	12	AAV15230	CD46 from clone pm
19	29	96.7	376	17	AAV93939	CD46 from clone pm
20	29	96.7	377	12	AAV15229	CD46 from clone pm
21	29	96.7	377	17	AAV93942	CD46 construct sub
22	29	96.7	377	17	AAV93941	CD46 construct sub
23	29	96.7	377	18	AAV27484	Human MCP. Homo s
24	29	96.7	384	12	AAV10924	Human MCP. Homo s
25	29	96.7	384	16	AAV86316	Human MCP. Homo s
26	29	96.7	418	21	AAV13589	Human MCP. Homo s
27	29	96.7	421	21	AAV58394	Human MCP. Homo s
28	29	96.7	421	22	AAV75528	Human MCP. Homo s
29	29	96.7	497	22	AAV04972	Human MCP. Homo s
30	29	96.7	520	22	AAV04739	Human MCP. Homo s
31	29	96.7	589	20	AAV34524	Human MCP. Homo s
32	29	96.7	590	21	AAV94928	Human MCP. Homo s
33	29	96.7	593	20	AAV34523	Human MCP. Homo s
34	29	96.7	606	20	AAV34393	Human MCP. Homo s
35	29	96.7	611	22	AAV03762	Human MCP. Homo s
36	29	96.7	612	22	AAV94498	Human MCP. Homo s
37	29	96.7	614	18	AAV20991	Human MCP. Homo s
38	29	96.7	621	22	AAV94049	Human MCP. Homo s
39	29	96.7	634	22	AAV40125	Human MCP. Homo s
40	29	96.7	634	22	AAV92742	Human MCP. Homo s
41	29	96.7	634	22	AAV83478	Human MCP. Homo s
42	29	96.7	634	22	AAV95120	Human MCP. Homo s
43	29	96.7	1464	19	AAV79294	Human MCP. Homo s
44	29	96.7	42	22	AAV15108	Human MCP. Homo s
45	29	96.7	42	22	AAV27563	Human MCP. Homo s

ALIGNMENTS

RESULT 1
ID AAV08336 standard: Peptide: 10 AA.

AAV08336:

19-JUL-1999 (first entry)

O. hannah venom alpha-neurotoxin homologous peptide 5.

Snake: venom: King cobra: alpha-neurotoxin: toxin: analgesia: screening.

Ophiophagus hannah.

MO9924055-A1.

20-MAY-1999.

03-NOV-1998: 98MO-SC00087.

06-NOV-1997: 97SC-0003972.

(UYST-) UNIV SINGAPORE NAT.

Gopaladrishtakone P, Gwee MCE, Kint RM, Pu XC, Wong PT:

WPT: 1999-327205/27.

Snake neurotoxin derived peptides

Claim 11: Page 43: 46pp: English.

This invention describes novel peptide derivatives of alpha-neurotoxin, derived from snake venom toxins of Ophiophagus hannah and which are

CC Capable of inducing or facilitating analgesia. The peptides, optionally
 CC with a carrier and/or diluent, are used to induce analgesia in a subject.
 CC Antibodies derived from the products of the invention can be used to
 CC screen for subject peptides, e.g. for purification purposes.
 XX
 XX
 SQ Sequence 10 AA;

Query Match 96.7%; Score 29; DB 20; Length 10;
 Best Local Similarity 80.0%; Pred. No. 13;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 11 11
 Db 3 fpcwr 7

RESULT 2

AAV08334 standard; Peptide: 11 AA.

XX AAV08334;

XX 19-JUL-1999 (first entry)

DE O. hannah venom alpha-neurotoxin homologous peptide 3.

XX Snake: venom; King cobra; alpha-neurotoxin; toxin; analgesia; screening.

OS Ophiophagus hannah.

XX WO9924055-A1.

XX 20-MAY-1999.

XX 03-NOV-1998; 98WO-SC00087.

XX 06-NOV-1997; 97SG-0003972.

XX (UYSI-) UNIV SINGAPORE NAT.

XX Gopaladrishtakone P, Gwee MCF, Kini RM, Pu XC, Wong PT;

XX WPI: 1999-327205/27.

XX Snake neurotoxin derived peptides

XX Claim 9; Page 42; 46pp; English.

XX This invention describes novel peptide derivatives of alpha-neurotoxin,
 CC derived from snake venom toxins of Ophiophagus hannah and which are
 CC capable of inducing or facilitating analgesia. The peptides, optionally
 CC with a carrier and/or diluent, are used to induce analgesia in a subject.
 CC Antibodies derived from the products of the invention can be used to
 CC screen for subject peptides, e.g. for purification purposes.
 XX
 XX
 SQ Sequence 11 AA;

Query Match 96.7%; Score 29; DB 20; Length 11;
 Best Local Similarity 80.0%; Pred. No. 14;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 11 11
 Db 3 fpcwr 7

RESULT 3

AAV58491 standard; Protein: 38 AA.

XX AAV58491;

XX 18-APR-1995 (first entry)

XX Binding domain of a polystyrene-binding TSAR (SB.9-5).

XX TSAR: totally synthetic affinity reagent; synthetic; binding domain;
 KW effector domain; concatenated heterofunctional protein; linker;
 KW direct; rapid; detection; screening; treatment; monoclonal antibody;
 KW Sm antigen; mouse model; autoimmune disease; SLE; MAD;
 KW systemic lupus erythematosus; polystyrene; SB.

OS Synthetic.

XX WO9418318-A.

XX 18-AUG-1994.

XX 01-FEB-1994; 94WO-US00977.

XX 01-FEB-1993; 93US-0013416.

XX 30-DEC-1993; 93US-0176500.

XX 31-JAN-1994; 94US-0189331.

XX (UYN-) UNIV NORTH CAROLINA.

XX Fowlkes DM, Kay BK;

XX WPI: 1994-279739/34.

XX Identifying proteins or peptide(s) which bind a ligand - by
 XX screening a recombinant vector library expressing fusion proteins
 XX comprising a binding domain and an effector domain

XX Example 7.8; Page 120; 255pp; English.

XX AAV58491-507 are amino acid sequences of the binding domain of TSAR
 CC (Totally Synthetic Affinity Reagents) peptides. These particular
 CC peptides are encoded by the polystyrene (SB) binding phage. There
 CC is no apparent linear motif, but the peptides are rich in Trp, Tyr and
 CC Gly, poor in Arg, Val and Lys, and completely lack cysteine residues.
 CC The non-variable amino acids at the N12 and COOH terminals are
 CC not shown. TSAR peptides are generated using generic oligonucleotides
 CC (see AA070470-73 for examples). TSARs are concatenated heterofunctional
 CC proteins or peptides, comprising at least two functional regions - a
 CC binding domain with affinity for a ligand and a second effector peptide
 CC portion that is chemically or biologically active. They may further
 CC comprise a linker peptide between the 2 domains. The TSARs or compns.
 CC comprising a TSAR binding domain can be used in vivo to deliver a
 CC chemically or biologically active moiety, eg. metal ion, radioisotope,
 CC can also replace the function of macromolecules eg. monoclonal or
 CC polyclonal antibodies and therefore circumvent the need for complex
 CC methods of hybridoma formation or in vivo antibody production. The TSARs
 CC are easily characterised and have designed activity allowing direct and
 CC rapid detection in a screening process.
 XX
 XX
 SQ Sequence 38 AA;

Query Match 96.7%; Score 29; DB 15; Length 38;
 Best Local Similarity 80.0%; Pred. No. 46;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 11 11
 Db 10 fpcwr 14

RESULT 4

AAV06599 standard; Protein: 94 AA.

XX AAV06599;

Oct-1999 (first entry)

Rat sodium channel NAN splice variant C-terminal sequence.

XX NaN; sodium channel; ion transport; rat; dorsal root ganglia;
KW pain; paresthesia; hyperexcitability; therapy.

XX Rattus sp.

XX MO9938889-A2.

XX 05-AUG-1999.

XX 29-JAN-1999: 99MO-US02008.

XX 20-NOV-1998: 98US-0109402.

XX 29-JAN-1998: 98US-0072990.

XX (UYVA) UNIV YALE.

XX Dlb-Hajj S, Waxman S;

XX WPI; 1999-479168/40.

XX New isolated nucleic acids encoding sodium channels, used to develop
PT products for treating acute or chronic pain or hyperexcitability
PT phenomena

XX Example 5; Page 35; 91pp; English.

XX This is the C-terminal sequence of a splice variant of rat sodium
XX channel NaN. This truncated variant comprises amino acids 1-1378
XX of the NaN sequence given in AAY06596 and the present, 94-amino acid
XX sequence at the C-terminal end. It thus lacks the 387 C-terminal
XX amino acids of full-length NaN. The variant arises from the use
XX of a cryptic donor splice site in exon 23 and a novel exon 23'
XX which is located in intron 23 of the rat NaN gene. NaN is a
XX previously unidentified voltage-gated sodium channel protein that
XX is preferentially expressed in dorsal root ganglia or trigeminal
XX ganglia, and which produces a TTX-R sodium current. The invention
XX provides methods for identifying agents that modulate NaN channel
XX activity or expression and for using such agents to treat acute or
XX chronic pain, paresthesia and hyperexcitability phenomena.

XX Sequence 94 AA:

QY 1 PPXMR 5 96.7%; Score 29; DB 20; Length 94;
1111
Db 52 fpawr 56 Best Local Similarity 80.0%; Pred. No. 1,1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 5

AAB20130 AAB20130 standard; Protein: 94 AA.

XX AAB20130;

XX 30-APR-2001 (first entry)

XX Rat truncated sodium channel C-terminal region.

XX Sodium channel; NaN; rat; tetrodotoxin resistant; pain;
KW paresthesia; hyperexcitability; analgesic; vaccine; therapy;
XX diagnosis.

XX Rattus norvegicus.

XX

PN WO200105831-A1.

XX 25-JAN-2001.

XX 14-JUL-2000; 2000MO-US19342.

XX 16-JUL-1999: 99US-0354147.

XX (UYVA) UNIV YALE.

XX Dlb-Hajj S, Waxman SG;

XX WPI; 2001-103147/11.

XX Nucleic acid molecules encoding human tetrodotoxin resistant sodium
PT channels, useful for preventing, diagnosing and treating pain,
PT paresthesia and/or hyperexcitability phenomena

XX Example 5; Page 34; 162pp; English.

XX The present sequence is that of the C-terminal portion of a
XX truncated rat NaN variant. The variant lacks the 387 C-terminal
XX amino acids of full-length rat NaN (see AAB20122), which are
XX replaced by the present 94-amino acid polypeptide. The N-terminal
XX 1378 amino acids of the truncated variant are identical to those
XX of full-length NaN. The new sequence arises from the use of a
XX cryptic donor splice site in exon 23 and a novel exon 23' which is
XX located in intron 23. Rat NaN is a novel tetrodotoxin resistant
XX sodium channel NaN belonging to the a-subunit voltage-gated sodium
XX channel protein family. These sodium channels underlie the
XX generation and propagation of impulses in excitable cells such as
XX neurons and muscle fibres. Preferential expression of NaN on
XX sensory neurons innervating the body (dorsal root ganglia) and
XX the face (trigeminal ganglia), but not on other neurons, makes
XX it a very useful target for diagnostic and/or therapeutic uses in
XX relation to acute and/or chronic pain pathologies. A claimed
XX method of treating pain, paresthesia and/or hyperexcitability
XX phenomena in a human or animal subject involves administering an
XX agent that alters sodium current flow through NaN channels, or
XX which modulates transcription or translation of NaN mRNA, in
XX dorsal root ganglia or trigeminal neurons. NaN polypeptides can
XX be obtained by recombinant expression, and used to treat disorders
XX associated with decreased sodium channel expression, to screen for
XX compounds that modulate sodium channel expression or activity,
XX and to raise antibodies useful as diagnostic agents.

XX Sequence 94 AA:

QY 1 PPXMR 5 96.7%; Score 29; DB 22; Length 94;
1111
Nb 52 fpawr 56 Best Local Similarity 80.0%; Pred. No. 1,1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 6

AAY04918 AAY04918 standard; Protein: 135 AA.

XX AAY04918;

XX 06-JUL-1999 (first entry)

XX Mycobacterium species protein sequence 35B.

XX Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KW hybridisation; detection; vaccine; immunisation; infection.

XX Mycobacterium sp.

XX

PN W09090186-A2.
 XX
 PD 25-FEB-1999.
 XX
 PF 14-AUG-1998: 98MO-FR01813.
 XX
 PR 11-SEP-1997: 97FR-0011325.
 PR 14-AUG-1997: 97FR-0010404.
 XX
 PA (INSP) INST PASTEUR.
 PI Gicquel, B., Lim EM, Pellicic V, Portnoi D, Goguet de la Salmoniere Y,
 PI Galignano A.
 DR WPI: 1999-181045/15.
 DR N-PSDB: AAX341171.
 XX
 PT Mycobacterial DNA vectors containing reporter constructs - for
 PT identifying coding or promoter sequences involved in
 PT infection-associated protein expression
 XX
 PS Claim 32: Fig 35B: 309pp: French.
 CC Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted
 CC proteins from various Mycobacterium species microorganisms. The
 CC encoding nucleotide sequences can be used as primers and probes for
 CC methods for detecting and identifying mycobacteria, especially belonging
 CC to the M. tuberculosis complex. The encoded proteins can be used in
 CC vaccines for immunisation against a bacterial or viral infection.
 CC
 SO Sequence 135 AA:

Query Match 96.7%: Score 29: DB 20: Length 135:
 Best Local Similarity 80.0%: Pred. No. 1.5e+02:
 Matches 4: Conservative 0: Mismatches 1: Indels 0: Gaps 0:
 QY 1 FPXMR 5
 11 11
 Db 66 fpawr 70

RESULT 7
 ID AAB58892 standard: Protein: 177 AA.
 XX
 AC AAB58892:
 XX
 DT 27-MAR-2001 (first entry)
 XX
 DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 600.
 XX
 KW Human: breast cancer; ovarian cancer; cytostatic; immunosuppressive;
 KW neotropic; neutroprotective; antiviral; antiallergic; hepatotropic;
 KW antidiabetic; antiinflammatory; antidiabetic; antidiabetic; antidiabetic;
 KW antidiabetic; antifungal; antiparasitic; cardiac; immune disorder;
 KW Addison's disease; allergy; autoimmune haemolytic anaemia;
 KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
 KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
 KW cardiovascular disorder; wound healing; neurological disease.
 XX
 OS Homo sapiens.
 XX
 PN W020005173-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000: 2000MO-US05881.
 XX
 PR 12-MAR-1999: 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX

PI Rosen CA, Ruben SM;
 XX
 DR WPI: 2000-611515/58.
 DR N-PSDB: AAF21795.
 XX
 PT New human breast and ovarian cancer associated gene sequences and the
 PT polypeptides encoded by these genes, useful in the prevention,
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular
 PT disorders and neurological diseases -
 XX
 PS Claim 11: Page 1037-1038: 1299pp: English.
 CC
 CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
 CC associated with breast and ovarian cancer. Included in the invention are
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
 CC isolation and characterisation of the DNA and protein sequences of the
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist
 CC or antagonist sequences exhibit cytostatic; immunosuppressive;
 CC neotropic; neutroprotective; antiviral; antiallergic; hepatotropic;
 CC antidiabetic; antiinflammatory; antidiabetic; antidiabetic; antidiabetic;
 CC antidiabetic; antifungal; antiparasitic and cardiac activity. The
 CC polynucleotide and protein sequences are used in the diagnosis of cancer,
 CC particularly breast and ovarian cancer. The nucleic acid sequences,
 CC proteins, agonists and antagonists may also be used in the diagnosis,
 CC prevention and treatment of immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; cardiovascular disorders such as
 CC myocardial ischaemia; wound healing; neurological diseases such as
 CC cerebral anoxia and epilepsy; and infectious diseases.
 CC
 SO Sequence 177 AA:

Query Match 96.7%: Score 29: DB 21: Length 177:
 Best Local Similarity 80.0%: Pred. No. 2e+02:
 Matches 4: Conservative 0: Mismatches 1: Indels 0: Gaps 0:
 QY 1 FPXMR 5
 11 11
 Db 22 fpswr 26

RESULT 8
 ID AAB52488 standard: Protein: 299 AA.
 XX
 AC AAB52488:
 XX
 DT 23-FEB-2001 (first entry)
 XX
 DE Helicobacter pylori bait polypeptide #6.
 XX
 KW Helicobacter pylori; two-hybrid system; protein-protein interaction;
 KW bait polypeptide; gastric ulcer; antibacterial.
 XX
 OS Helicobacter pylori.
 XX
 PN W0200066722-A1.
 XX
 PD 09-NOV-2000.
 XX
 PF 14-APR-2000: 2000MO-IB00603.
 XX
 PR 30-APR-1999: 99EP-0401066.
 XX
 PA (HYBR-) HYBRIGENICS SA.
 XX
 PI Legrain P, Seling L, Rain J;
 XX
 DR WPI: 2000-687535/67.
 DR N-PSDB: AAC97234.

PT A two-hybrid system for identifying compounds useful in the treatment
 of e.g. gastric ulcers comprises producing a collection of recombinant
 cell clones -

PS Example 5: Page 88-89; 267pp: English.

CC the present sequence is a bait polypeptide used in a Helicobacter
 CC pylori two-hybrid screen to identify protein-protein interactions.
 CC The method is used to identify a recombinant cell clone expressing a
 CC prey polypeptide which is capable of interacting with the bait
 CC polypeptide. The two hybrid system is useful for screening compounds
 CC for antibacterial activity. It may be used in the treatment of gastric
 CC ulcers. The polynucleotides are useful as amplification primers or
 CC specific detection probes. The polypeptides, vectors or host cells can
 CC be used as immunogens to produce mono- or polyclonal antibodies. The
 CC polynucleotides, polypeptides, antibodies, vectors, host cells or
 CC modulating agents can be used to produce a pharmaceutical composition.

CC Sequence 299 AA:

Query Match 96.7%; Score 29; DB 21; Length 299;

Best Local Similarity 80.0%; Pred. No. 3.3e+02; Mismatches 1; Indels 0; Gaps 0;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5

Db 89 fpswr 93

RESULT 9

ID AAR93943 standard; Protein: 314 AA.

AC AAR93943;

DT 21-MAY-1996 (first entry)

DE CD46 construct delSCR2/subSCR3.

XX CD46: recombinant protein; short consensus repeat; SCR:

KW regulator of complement activation; transgenic animal; pig;

KW organ transplantation.

XX Synthetic.

XX WO9606937-A1.

PD 07-MAR-1996.

PF 30-AUG-1995; 95WO-AU00553.

PR 30-AUG-1994; 94AU-0007724.

PA (AUST-) AUSTIN RES INST.

PI Christiansen D, Loveland B, McKenzie JFC, Milland J;

DR WPI: 1996-160368/16.

DR N-PSDB: AAT17599.

PT Increasing prodn. of recombinant proteins, esp. CD46 - by reducing

PT the amt. of A and/or T in an A and/or T rich region of encoding gene

PT exon

PS Claim 12: Page 40-41; 60pp: English.

CC CD46 delSCR2/subSCR3 (AAR93943) is the product of a cDNA construct

CC (AAT17599) obtd. by splice overlap extension PCR of wild-type CD46

CC cDNA (AAT17595). The A+T content of A+T-rich exon 5 of the gene,

CC encoding the short consensus repeat (SCR) 3 of CD46, was lowered

CC and SCR2 was deleted. This resulted in improved prodn. in.

CC eukaryotic host cells, e.g. CHO-K1 and COS-7. CD46 delSCR2/subSCR3
 CC is used to prevent complement- or inflammation-mediated tissue
 CC damage, to improve immunity to tumours or viruses, to control
 CC fertilisation and to prevent spontaneous abortion. Expression in
 CC transgenic animals, esp. pigs, provides organs suitable for
 CC transplantation.

SO Sequence 314 AA:

Query Match 96.7%; Score 29; DB 17; Length 314;

Best Local Similarity 80.0%; Pred. No. 3.4e+02; Mismatches 1; Indels 0; Gaps 0;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5

Db 11 fpswr 15

RESULT 10

ID AAR15233 standard; Protein: 324 AA.

AC AAR15233;

DT 17-MAR-1992 (first entry)

DE CD46 from clone pm5.8.

XX SCR: short consensus repeats; transmembrane; cytoplasmic;

KW membrane cofactor protein; MCP.

XX Homo sapiens.

XX Key

FT Peptide 1..34

FT Protein 34..324

FT Modified-site 83

FT Modified-site 114

FT Domain 289..304

XX WO9118097-A.

XX 28-NOV-1991.

PF 10-MAY-1991; 91WO-AU00199.

PR 11-MAY-1990; 90AU-0000133.

PA (UWME-) UNIV MELBOURNE.

PI Purcell DJJ, Russell SM, McKenzie JFC;

DR WPI: 1991-369251/50.

DR N-PSDB: AAQ14919.

PT New CD46 membrane co-factor protein variants - useful as probes

PT to identify CD46 isoforms and for diagnosing spontaneous

PT abortion, inhibiting immune response and treating leukaemia

PS Disclosure: Fig 1A and 3B; 77pp: English.

CC The sequence of the pm5.8 clone is identical to the pm5.1 clone in

CC the portion encoding the N12 leader and four SCR regions. The

CC sequence after nucleotide 890 was different, and results from

CC reading through of the cDNA into an intron sequence after the

CC fourth SCR. The protein encoded by this new sequence encodes

CC a 16 amino acid hydrophobic region, probably serving as a membrane

CC spanning sequence. The putative cytoplasmic tail would also
 CC contain a high proportion of charged amino acids.
 CC See also AAQ14915-25, AAQ15211-12 and AAR15457-59.
 XX
 SO Sequence 324 AA;

Query Match 96.7%; Score 29; DB 12; Length 324;
 Best Local Similarity 80.0%; Pred. NO. 3.5e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 11 11
 Db 11 fpswr 15

RESULT 11

AAP70396
 ID AAP70396 standard; Protein: 359 AA.

XX
 AC AAP70396;

DT 30-APR-1991 (first entry)

DE Cellulase.

KM Cellulase: plasmid; transformation.

OS Cellulomonas uda CB4 (FER.

XX JP62175178-A.

PN 31-JUL-1987.

XX 30-JAN-1986; 86JP-0016800.

PE 30-JAN-1986; 86JP-0016800.

XX (ENER-) SHIN ENERGY SOGO KAIHATSU KIKO.

PA (NENR-) NENRYOYO ALCOHOL KAIHATSU CLUTSU KENKYU KUMIAT.

XX WPI: 1987-253827/36.

DR N-PSDB: AAN70617.

XX Bio-engineering cellulase prodn. - by preparing DNA chain,
 PT preparing plasmid conty. DNA chain in the gene expressible state,
 PT and transforming bacterial by plasmid

XX Disclosure: Fig 1(A-B): 12pp; Japanese.

CC Transformation E.coli JM103 (FERM P-8612) with the gene encoding
 CC this protein allows expression of the protein and easy and cheap
 CC prodn. of cellulase.

XX Sequence 359 AA;

SO Sequence 359 AA;

Query Match 96.7%; Score 29; DB 8; Length 359;
 Best Local Similarity 80.0%; Pred. NO. 3.9e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 11 11
 Db 178 fpswr 182

RESULT 12

AAR27793
 ID AAR27793 standard; Protein: 372 AA.

XX
 AC AAR27793;

DT 12-MAR-1993 (first entry)

XX New platelet factor 4 receptor superfamily member PF4AR11.
 DE
 XX
 KW IL-8R; G-protein coupled receptor family; rhodopsin superfamily;
 KW pro-inflammatory cytokine; Bfr.9.
 XX

OS Homo sapiens.

XX W09217497-A.

PN 15-OCT-1992.

XX 23-MAR-1992; 92MO-US02317.

PR 29-MAR-1991; 91US-0677211.

XX 19-DEC-1991; 91US-0810782.

PA (GETH) GENENTECH INC.

XX Holmes WE, Lee J, Wood WI;

DR WPI: 1992-366191/44.

XX N-PSDB: AAQ37107.

XX Isolated human platelet factor 4 super-family receptor
 PT polypeptide and corresp. antibodies and DNA - useful as
 PT diagnostic and screening agents, and for treating inflammation or
 PT PF4AR-mediated disorders

XX Claim 7; Fig 5; 78pp; English.

XX The IL-8 receptor cDNA sequence was isolated (see AAQ29505) and a
 CC 874bp sub-fragment of the coding sequence was used as a probe to
 CC screen human cell line HU60 and human peripheral blood lymphocyte
 CC cDNA libraries. Two new gene sequences were found that are clearly
 CC related to the IL-8 receptor. One of these was contained in clone
 CC Bfr.9 and is predicted to encode an amino acid sequence which is
 CC 36% and 38% identical with the high and low affinity IL-8 receptor
 CC sequences, respectively. See also AAQ37107.

XX Sequence 372 AA;

SO Sequence 372 AA;

Query Match 96.7%; Score 29; DB 13; Length 372;
 Best Local Similarity 80.0%; Pred. NO. 4e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 11 11
 Db 352 fpswr 356

Query Match 96.7%; Score 29; DB 13; Length 372;
 Best Local Similarity 80.0%; Pred. NO. 4e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 11 11
 Db 352 fpswr 356

RESULT 13
 ID AAR92239 standard; Protein: 372 AA.

XX
 AC AAR92239;

DT 26-MAR-1996 (first entry)

XX Chemokine superfamily receptor.

DE Interleukin; IL-8; inflammation; psoriasis; dermatitis;
 KW rheumatoid arthritis; inflammatory bowel disease;
 KW chronic lung inflammation; treatment; antibody;
 KW affinity purification; detection.

XX Homo sapiens.

OS US5440021-A.

XX 08-AUG-1995.

PD 08-AUG-1995.

XX

Query Match	96.7%	Score 29:	DB 16:	Length 372:
Best Local Similarity	80.0%	Pred No.	4e+02:	
Matches 4:	Conservative	0:	Mismatches 1:	Indels 0:
				Caps 0

XX	WIPI: 1995-036114/005.
DR	N-PSDB; AAQ80522.
XX	Treatment of inflammatory disorders - by administering an antibody capable of binding a platelet factor 4 superfamily receptor polypeptide
PT	
XX	
PS	Disclosure: Page 56-58; 83pp; English.
XX	
CC	2 PF4AR members were identified by probing lambda libraries from human monocytic-like cell line HL-60 and human peripheral blood lymphocytes using a large fragment of IL-8 receptor DNA (full sequence given in AAQ80520). The nucleotide sequences of the 2 PF4ARs are given in AAQ80521 and AAQ80522, and their respective amino acid sequences in AAR68812 and AAR68813.
CC	
XX	
SQ	Sequence 372 AA:
Gy	1 PPXMR 5
	1111
Db	352 fpswr 356
RESULT 15	
AY006644	
ID	AAV06644 standard; Protein: 372 AA.
XX	
AC	AAV06644;
XX	
DIF	26-OCT-1999 (first entry)
XX	
LDE	Human Burkitt's lymphoma receptor 1 (BLR1).
XX	
KW	Burkitt's lymphoma receptor 1; BLR1; human; B lymphocyte chemoattractant; BLC; chemokine; ligand; drug screening; leukemia; autoimmune disease; therapy.
KW	
XX	
OS	Homo sapiens.
XX	
PN	WO9928468-A1.
XX	
PD	10-JUN-1999.
XX	
PF	02-DEC-1998; 98WO-US25561.
XX	
PR	02-DEC-1997; 97US-0982493.
XX	
PA	(RECC) UNIV CALIFORNIA.
XX	
PI	Cyster JG, Gunn MD, Williams LT;
XX	
JR	WPI: 1999-493764/41.
DR	N-PSDB; AAX87710.
XX	
PT	Modulating interaction of a Burkitt's Lymphoma Receptor 1 polypeptide and ligand, useful in drug screens
XX	
XX	
PS	Claim 1: Page 38; 42pp; English.
XX	
CC	This sequence represents human Burkitt's lymphoma receptor 1 (BLR1). The invention relates to methods for modulating the interaction of BLR1 with its ligand, a lymphocyte chemoattractant (BLC, see AAV06642). The methods comprise combining BLR1 and BLC polypeptides with a candidate modulator agent under conditions whereby, but for the presence of the agent, the polypeptides engage in a first interaction, and determining a second interaction of the polypeptides in the presence of the agent, wherein a difference

CC between the first and second interactions indicates that the agent
CC modulates the interaction of the polypeptides. The modulator is
CC preferably an antagonist, especially dominant negative, form of
CC BLC. BLC and BLC agonists and antagonists may be useful for
CC treating viral (e.g. HIV) infections, lymphoma, B lineage
CC leukemia, and autoimmune diseases such as rheumatoid arthritis,
CC thyroiditis and diabetes.

XX
SQ Sequence 372 AA;

Query Match: 96.7%; Score 29; DB 20; Length 372;

Best Local Similarity 80.0%; Pred. No. 4e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5

DB 352 fpxmr 356

Search completed: February 27, 2002, 11:41:19
Job time: 458 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:45:22 : Search time 145.23 seconds
(without alignments)
2.623 Million cell updates/sec

Title: US-09-446-109a-14
Perfect score: 30
Sequence: 1 FPXMR 5

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database: PIR-68:
1: pirt:
2: pirt:
3: pirt:
4: pirt:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	96.7	73	1 N20H1	long neurotoxin 1
2	29	96.7	113	2 T07855	translational elonga
3	29	96.7	157	2 T24380	hypothetical prote
4	29	96.7	245	2 D82680	conserved hypothet
5	29	96.7	256	2 D85510	probable EC 3.5. a
6	29	96.7	259	2 A29831	heat-labile entero
7	29	96.7	273	2 T05454	hypothetical prote
8	29	96.7	292	2 G82960	hypothetical prote
9	29	96.7	327	2 S56162	MDCKR15 protein - h
10	29	96.7	332	2 JT0585	minor endoglucanas
11	29	96.7	349	2 G02913	sperm CD46 - human
12	29	96.7	359	2 I40696	endoglucanase - Ce
13	29	96.7	369	2 I57998	membrane cofactor
14	29	96.7	372	2 S26667	G protein-coupled
15	29	96.7	377	2 I54479	membrane cofactor
16	29	96.7	384	2 S01896	protein F25C20.9 (
17	29	96.7	417	2 D86251	aldehyde dehydroge
18	29	96.7	479	2 T35441	probable aldehyde
19	29	96.7	497	2 D83628	heterodisulfide re
20	29	96.7	555	2 F69312	hypothetical prote
21	29	96.7	625	2 H70330	ferrous iron trans
22	29	96.7	642	2 D71909	iron(II) transport
23	29	96.7	650	2 T22002	hypothetical prote
24	29	96.7	735	2 A83006	hypothetical prote
25	29	96.7	735	2 T40325	dermonecrotic tox1
26	29	96.7	735	2 T40325	transposase ssr289
27	29	96.7	735	2 T40325	hypothetical prote
28	29	96.7	735	2 T40325	hypothetical prote
29	29	96.7	735	2 T40325	hypothetical prote

30	28	93.3	93	2 S75008	transposase ss192
31	28	93.3	103	2 T47718	hypothetical prote
32	28	93.3	114	2 S77061	transposase sl1066
33	28	93.3	143	1 HKK48	heat shock protein
34	28	93.3	143	1 HKK41	heat shock protein
35	28	93.3	147	2 A71560	hypothetical prote
36	28	93.3	149	2 T35846	probable integral
37	28	93.3	157	2 F81710	conserved hypothet
38	28	93.3	169	2 G69300	conserved hypothet
39	28	93.3	254	2 S76459	transposase sl1186
40	28	93.3	254	2 E82791	conserved hypothet
41	28	93.3	258	2 S45682	acetone-cyanhydrin
42	28	93.3	259	2 S76643	transposase slr051
43	28	93.3	261	2 S77171	transposase sl1171
44	28	93.3	261	2 S77351	transposase sl1171
45	28	93.3	261	2 S75081	transposase slr026

ALIGNMENTS

RESULT 1
N20H1
Long neurotoxin 1 - king cobra
N:Alternate names: neurotoxin A
C:Species: Ophiophagus hannah (king cobra)
C:Date: 22-Jun-1981 #sequence_revision 22-Jun-1981 #text_change 23-Aug-1996
C:Accession: A01658
R:Joubert, F.J.
Biochim. Biophys. Acta 317, 85-98, 1973
A:Title: Snake venom toxins. The amino acid sequences of two toxins from Ophiophagus
A:Reference number: A90588; PMID:7321298
A:Accession: A01658
A:Molecule type: protein
A:Residues: 1-73 <JUN>
C:Superfamily: Snake toxin
C:Keywords: neurotoxin; venom
F:3-21,14-42,27-31,46-57,58-63/Disulfide bonds: #status predicted

Query Match 96.7% Score 29; DB 1; Length 73;
Best Local Similarity 80.0%; Pred. No. 33;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 FPXMR 5
UB 66 FPTWR 70

RESULT 2
N20H1
translational elongation factor eEF-1 alpha chain - rape (fragment)
N:Alternate names: translation elongation factor EF-1A
C:Species: Brassica napus (rape)
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 31-Mar-2000
C:Accession: T07855
K:Saez-Vasquez, J.
submitted to the EMBL Data Library, February 1995
A:Reference number: Z16177
A:Accession: T07855
A:Status: preliminary; translated from GH/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-113 <SAE>
A:Cross-references: EMBL:021744; NID:g914912; PIDN:AA86366.1; PID:g914913
A:Experimental source: cv. Samoural; 5 days old etiolated seedlings
C:Superfamily: translation elongation factor Tu; translation elongation factor Tu hom
C:Keywords: GTP binding; protein biosynthesis

Query Match 96.7% Score 29; DB 2; Length 113;
Best Local Similarity 80.0%; Pred. No. 50;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
 Db 103 FPMWR 107

RESULT 3

T24380
 hypothetical protein T03D8.2 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T24380
 R:Morimoto, D.
 submitted to the EMBL Data Library, March 1997
 A:Reference number: Z19884
 A:Accession: T24380
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1157 <MIL>
 A:Cross-references: EMBL:Z292838; PIDN:CAB07406.1; GSPDB:GN00023; CESP:T03D8.2
 A:Experimental source: clone T03D8
 C:Genetics:
 A:Gene: CESP:T03D8.2
 A:Map position: 5
 A:Introns: 49/3; 83/1; 125/3

Query Match 96.7%: Score 29; DB 2; Length 157;
 Best Local Similarity 80.0%: Pred. No. 68;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
 Db 40 FPMWR 44

RESULT 4

D82880
 conserved hypothetical protein XF1438 [imported] - *Xylella fastidiosa* (strain 9a5c)
 C:Species: *Xylella fastidiosa*
 C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: D82880
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
 Nature 406, 151-157, 2000
 A>Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
 A:Reference number: A82515; MUID:20365717
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: D82880
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-245 <SIM>
 A:Cross-references: GB:AE003974; GB:AE003849; NID:91106454; PIDN:AAF84247.1; GSPDB:GN001
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
 Brites, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir, I
 as-Neto, E.; Docena, C.; El-Dorfy, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to Genbank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; From
 J.D.; Junqueira, M.L.; Kemper, F.L.; Kitejima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
 chdo, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins, H
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
 M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XF1438

Query Match 96.7%: Score 29; DB 2; Length 245;
 Best Local Similarity 80.0%: Pred. No. 1e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
 Db 197 FPMWR 201

RESULT 5

D85510
 Probable EC 3.5. amidase-type enzyme yafV [imported] - *Escherichia coli* (strain O157:
 C:Species: *Escherichia coli*
 C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
 C:Accession: D85510
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
 Jiler, L.; Grobeck, E.J.; Davis, N.W.; Linn, A.; Dimalanta, E.; Potamousis, K.; Apoda
 Nature 409, 529-533, 2001
 A>Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: D85510
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-256 <STO>
 A:Cross-references: GB:AE005174; NID:912512987; PIDN:AAC54544.1; GSPDB:GN00145; UNCP:
 A:Experimental source: strain O157:H7, Substrain EDL933
 C:Genetics:
 A:Gene: yafV

Query Match 96.7%: Score 29; DB 2; Length 256;
 Best Local Similarity 80.0%: Pred. No. 1e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
 Db 244 FPMWR 248

RESULT 6

A29831
 heat-labile enterotoxin IIA chain A precursor - *Escherichia coli*
 N:Alternate names: LT-IIa
 C:Species: *Escherichia coli*
 C>Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
 C:Accession: A29831
 R:Pickett, C.L.; Weinstein, D.L.; Holmes, R.K.
 J. Bacteriol. 169, 5180-5187, 1987
 A>Title: Genetics of type IIA heat-labile enterotoxin of *Escherichia coli*: operon fus
 A:Reference number: A91849; MUID:88032841
 A:Accession: A29831
 A:Molecule type: DNA
 A:Residues: 1-259 <PIC>
 A:Cross-references: GB:M17894; NID:9146671; PIDN:AAA4093.1; PID:9146672
 A:Note: the authors translated the codon TAT for residue 225 as Thr
 C:Superfamily: heat-labile enterotoxin chain A
 F:1-18/Domain: signal sequence #status predicted <SIG>
 F:19-259/Product: heat-labile enterotoxin IIA chain A #status predicted <LTA>

Query Match 96.7%: Score 29; DB 2; Length 259;
 Best Local Similarity 80.0%: Pred. No. 1e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
 Db 187 FPMWR 191

RESULT 7

T05454
 hypothetical protein F7K2.180 - *Arabidopsis thaliana*
 C:Species: *Arabidopsis thaliana* (mouse-ear cross)
 C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Sep-1999
 C:Accession: T05454
 R:Bevan, M.; Medler, H.; Wambolt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schue

submitted to the Protein Sequence Database, November 1998

A:Reference number: 215416
A:Accession: T05454
A:Molecule type: DNA
A:Residues: 1-273 <BEV>
A:Cross-references: EMBL:AL033545
A:Experimental source: Cultivar Columbia; BAC clone F7K2
C:Genetics:
A:Map position: 4
A:Note: F7K2.180
C:Superfamily: Arabidopsis hypothetical protein F7K2.180

Query Match 96.7%; Score 29; DB 2; Length 273;
Best Local Similarity 80.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
DB 79 FPSWR 83

RESULT 8
G82960
hypothetical protein PA5488 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: G82960
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Jarbig, K.; Lim,
J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950, MUID:20437337
A:Accession: G82960
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-292 <STO>
A:Cross-references: GB:AE004961; GB:AE004091; NID:g9951814; PIDN:AAC08873.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA5488

Query Match 96.7%; Score 29; DB 2; Length 292;
Best Local Similarity 80.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
DB 231 FPSWR 235

RESULT 9
S56162
MCCR15 protein - human
C:Species: Homo sapiens (man)
C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999
C:Accession: S56162
R:Barclay, L.; Loetscher, M.; Tobler, A.; Baggiolini, M.; Moser, B.
Blochem. J. 309, 773-779, 1995
A:Title: Sequence variation of a novel heptahelical leucocyte receptor through alternat
A:Reference number: S56162; MUID:95366951
A:Accession: S56162
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-327 <BAR>
A:Cross-references: EMBL:X68829; NID:g840783; PIDN:CAA48723.1; PID:g840784
C:Superfamily: vertebrate rhodopsin

Query Match 96.7%; Score 29; DB 2; Length 327;
Best Local Similarity 80.0%; Pred. No. 1.4e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 FPXWR 5
DB 307 FPSWR 311

RESULT 10
J70585
minor endoglucanase (EC 3.2.1.-) precursor - Erwinia chrysanthemi
C:Species: Erwinia chrysanthemi
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 08-Oct-1999
C:Accession: J70585; J70586
R:Guiseppi, A.; Aymeric, J.L.; Caml, B.; Barras, F.; Creuzet, N.
Gene 106, 109-114, 1991
A:Title: Sequence analysis of the cellulase-encoding celY gene of Erwinia chrysanthem
A:Reference number: J70585; MUID:92039050
A:Accession: J70585
A:Molecule type: DNA
A:Residues: 1-332 <GU1>
A:Cross-references: GB:M74044; NID:g148391; PIDN:AAZ4818.1; PID:g148392
A:Experimental source: strain 3937
A:Accession: J70586
A:Molecule type: protein
A:Residues: 24-33 <GU11>
C:Genetics:
A:Gene: celY
C:Keywords: glycosidase; hydrolase
F:1-23/Domain: signal sequence #status predicted <STC>
F:24-332/Product: minor endoglucanase #status predicted <MIN>

Query Match 96.7%; Score 29; DB 2; Length 332;
Best Local Similarity 80.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
DB 178 FPSWR 182

RESULT 11
G02913
sperm CD46 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 09-Jun-2000
C:Accession: G02913
R:Hara, T.
submitted to GenBank, March 1996
A:Reference number: H01942
A:Accession: G02913
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-349 <HAR>
A:Cross-references: GB:D84105; NID:g1256700; PIDN:BA12224.1; PID:g1256701
C:Superfamily: herpesvirus complement control protein; complement factor H repeat hom
F:35-94/Domain: complement factor H repeat homology <FH01>
F:99-157/Domain: complement factor H repeat homology <FH02>
F:162-223/Domain: complement factor H repeat homology <FH03>
F:228-283/Domain: complement factor H repeat homology <FH04>

Query Match 96.7%; Score 29; DB 2; Length 349;
Best Local Similarity 80.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
DB 11 FPSWR 15

RESULT 12
140696

endoglucanase - Cellulomonas uda
 C:Species: Cellulomonas uda
 C:Date: 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 05-Nov-1999
 C:Accession: I40696
 R:Nakamura, K.; Misawa, N.; Kitamura, K.
 J. Biotechnol. 4: 247-254, 1986
 A:Title: Sequence of a cellulase gene of Cellulomonas uda CB4.
 A:Reference number: 140696
 A:Accession: I40696
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-359 <RES>
 A:Cross-references: GB:M36503; NID:g144430; PIDN:AAA23090.1; PID:g144431

Query Match 96.7%; Score 29; DB 2; Length 359;
 Best Local Similarity 80.0%; Pred. No. 1.5e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 II II
 DB 178 FPMWR 182

RESULT 13
 157998
 membrane cofactor protein - human
 C:Species: Homo sapiens (man)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 19-May-2000
 C:Accession: I57998
 R:Cervoni, F.; Feniichel, P.; Akhoundi, C.; Hsi, B.L.; Rossi, B.
 Mol. Reprod. Dev. 34, 107-113, 1993
 A:Title: Characterization of a cDNA clone coding for human testis membrane cofactor protein
 A:Reference number: I57998; MUID:93119658
 A:Accession: I57998
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-369 <RES>
 A:Cross-references: GB:S51940; NID:q262937; PIDN:AA824802.1; PID:q262938
 C:Superfamily: herpesvirus complement control protein; complement factor H repeat homolo
 F:35-94/Domain: complement factor H repeat homology <FH01>
 F:99-157/Domain: complement factor H repeat homology <FH02>
 F:162-223/Domain: complement factor H repeat homology <FH03>
 F:228-283/Domain: complement factor H repeat homology <FH04>

Query Match 96.7%; Score 29; DB 2; Length 369;
 Best Local Similarity 80.0%; Pred. No. 1.5e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 II II
 DB 11 FPMWR 15

RESULT 14
 S26667
 G protein-coupled receptor BURL - human
 C:Species: Homo sapiens (man)
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jul-2000
 C:Accession: S26667
 R:Dobner, T.; Wolf, I.; Emrich, T.; Lipp, M.
 Eur. J. Immunol. 22, 2795-2799, 1992
 A:Title: Differential expression of a novel G protein-coupled receptor from
 A:Reference number: S26667; MUID:93049615
 A:Accession: S26667
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-372 <DOB>
 A:Cross-references: EMBL:X68149; NID:g29459; PIDN:CAA48252.1; PID:g29460
 C:Genetics:
 A:Gene: GDB:BURL
 A:Cross-references: GDB:136235; OMIM:601613

A:Map position: 15q26.1-15q26.1
 C:Superfamily: vertebrate rhodopsin
 C:Keywords: G protein-coupled receptor; transmembrane protein

Query Match 96.7%; Score 29; DB 2; Length 372;
 Best Local Similarity 80.0%; Pred. No. 1.5e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 II II
 DB 352 FPMWR 356

RESULT 15
 I54479
 membrane cofactor protein precursor, splice form pm5.1 - human
 N:Alternate names: lymphocyte surface glycoprotein CD46
 C:Species: Homo sapiens (man)
 C:Date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 31-Mar-2000
 C:Accession: I54479
 R:Putcell, D.F.; Russell, S.M.; Deacon, N.J.; Brown, M.A.; Hooker, D.J.; McKenzie, I.
 Immunogenetics 33, 335-344, 1991
 A:Title: Alternatively spliced RNAs encode several isoforms of CD46 (MCP), a regulator
 A:Reference number: I54479; MUID:91267562
 A:Accession: I54479
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-377 <RES>
 A:Cross-references: GB:M58050; NID:g180136; PIDN:AAA62833.1; PID:g180137
 C:Genetics:
 A:Gene: GDB:MCP
 A:Cross-references: GDB:120169; OMIM:120920
 A:Map position: 1q32-1q32
 A:Function:
 C:Description: for the factor I-mediated cleavage of the complement convertases
 C:Superfamily: herpesvirus complement control protein; complement factor H repeat hom
 C:Keywords: alternative splicing; glycoprotein; transmembrane protein
 F:35-94/Domain: complement factor H repeat homology <FH01>
 F:99-157/Domain: complement factor H repeat homology <FH02>
 F:162-223/Domain: complement factor H repeat homology <FH03>
 F:228-283/Domain: complement factor H repeat homology <FH04>
 F:329-351/Domain: transmembrane #status predicted <TM>

Query Match 96.7%; Score 29; DB 2; Length 377;
 Best Local Similarity 80.0%; Pred. No. 1.6e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 II II
 DB 11 FPMWR 15

Search completed: February 27, 2002, 11:45:23
 Job time: 702 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OK protein - protein search, using sw model

Run on: February 27, 2002, 11:42:48 ; Search time 78.39 Seconds
(without alignments)
2.339 Million cell updates/sec

Title: US-09-446-109A-14

Perfect score: 30
Sequence: 1 FPXMR 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	96.7	73	1	NKX1.0PHNA
2	29	96.7	165	1	DSBB_VIBAL
3	29	96.7	259	1	E2AA_ECOLI
4	29	96.7	332	1	GUNY_ERMCH
5	29	96.7	359	1	GUNY_CELUD
6	29	96.7	372	1	CCR5_HUMAN
7	29	96.7	377	1	MCP_HUMAN
8	29	96.7	420	1	2229_HUMAN
9	28	93.3	143	1	HS16_CAEEL
10	28	93.3	143	1	HS17_CAEEL
11	28	93.3	257	1	HNL_MANES
12	28	93.3	271	1	PRBC_VIBRA
13	28	93.3	271	1	PRBC_BACST
14	28	93.3	286	1	PAL_KLEPN
15	28	93.3	289	1	PAL_ECOLI
16	28	93.3	289	1	PAL_PROVU
17	28	93.3	289	1	PAL_SALTY
18	28	93.3	298	1	CC28_YEAST
19	28	93.3	332	1	CC2_CAEEL
20	28	93.3	374	1	CCR5_MOUSE
21	28	93.3	374	1	CCR5_RAT
22	28	93.3	430	1	KOTA_CHLMU
23	28	93.3	431	1	KOTA_CHLTR
24	28	93.3	433	1	ANM2_HUMAN
25	28	93.3	437	1	KOTA_CHLPN
26	28	93.3	448	1	Y347_CHLPN
27	28	93.3	459	1	RBL2_THIDE
28	28	93.3	463	1	RBL3_HYDMR
29	28	93.3	470	1	LTP2_MYOCO
30	28	93.3	482	1	CATTA_ONCYE
31	28	93.3	490	1	CE05_ECOLI
32	28	93.3	490	1	CE10_ECOLI
33	28	93.3	492	1	CAT2_CUCPE

34	28	93.3	1056	1	CAT3_CUCPE
35	28	93.3	766	1	CEAK_ECOLI
36	28	93.3	902	1	EYA_DROME
37	28	93.3	973	1	FTDH_RAT
38	28	93.3	973	1	TRP5_HUMAN
39	28	93.3	974	1	TRP4_MOUSE
40	28	93.3	974	1	TRP5_RABIT
41	28	93.3	975	1	TRP5_MOUSE
42	28	93.3	977	1	TRP4_HUMAN
43	28	93.3	977	1	TRP4_RAT
44	28	93.3	981	1	TRP4_BOVIN
45	28	93.3	1056	1	DPOL_ADE02

ALIGNMENTS

RESULT 1	hX1.0PHNA	STANDARD:	PRT: 73 AA.
ID	NKX1.0PHNA		
AC	P01387:		
UT	21-JUL-1986 (Rel. 01, Created)		
DT	21-JUL-1986 (Rel. 01, Last sequence update)		
BT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	LONG NEUROTOXIN 1 (NEUROTOXIN A).		
OS	Ophiophagus hannah (King cobra)		
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidae;		
CC	Elapidae; Elapinae; Ophiophagus.		
OX	NCBI_TaxID=8665;		
IN	[1]		
RP	SEQUENCE:		
RC	TISSUE=Venom;		
RX	MEDLINE=73231298; PubMed=4198767;		
FA	Joubert F.J.:		
FT	"Snake venom toxins the amino acid sequences of two toxins from		
FT	Ophiophagus hannah (King cobra) venom."		
FL	Biochim. Biophys. Acta 317:85-98(1973).		
CC	-1- SUBCELLULAR LOCATION: SECRETED.		
CC	-1- TISSUE SPECIFICITY: EXPRESSED BY VENOM GLAND.		
CC	-1- MISCELLANEOUS: LD(50) IS 0.3 MG/KG BY SUBCUTANEOUS INJECTION.		
CC	-1- SIMILARITY: BELONGS TO THE SNAKE TOXIN FAMILY.		
DR	PIR: A01658; N20H1.		
DR	HSSP: P01386; ITXB.		
DR	Interpro: IPR003571; Snake-toxin.		
DR	Pfam: PF00087; Toxin: 1.		
DR	ProDom: PD000206; Snake-toxin: 1.		
DR	PROSITE: PS00272; SNAKE_TOXIN: 1.		
KW	Venom; Neurotoxin; Multigene family.		
FT	DISULFID 3 21 BY SIMILARITY.		
FT	DISULFID 14 42 BY SIMILARITY.		
FT	DISULFID 27 31 BY SIMILARITY.		
FT	DISULFID 46 57 BY SIMILARITY.		
FT	DISULFID 58 63 BY SIMILARITY.		
SO	SEQUENCE 73 AA: 8106 MW: 1AC17E91E16C54F7 CRC64:		

Query Match 96.7% Score 29; DB 1; Length 73;
Best Local Similarity 80.0%; Pred. NO. 17;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5

ND 66 FETWR 70

RESULT 2

DSBB_VIBAL

STANDARD:

PRT: 165 AA.

AC Q56578:

BT 01-NOV-1997 (Rel. 35, Created)

BT 01-NOV-1997 (Rel. 35, Last sequence update)

BT 20-AUG-2001 (Rel. 40, Last annotation update)

DE DISULFIDE BOND FORMATION PROTEIN B (DISULFIDE OXIDOREDUCTASE)
 DE (FRAGMENT).
 GN DSB.
 OS Vibrio alginolyticus.
 OC Bacteria: Proteobacteria: gamma subdivision: Vibrionaceae: Vibrio.
 OX NCBI_TaxID=663;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-B138-2;
 RA Nakamura T., Enomoto H., Unemoto T.;
 RL Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: REQUIRED FOR DISULFIDE BOND FORMATION IN SOME
 CC PERIPLASMIC PROTEINS. ACTS BY OXIDIZING THE DSB. PROTEIN (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
 CC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE DSB. FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D83728; BAA12087.1; -
 DR InterPro: IPR003752; DSB.
 DR Pfam: PF02600; DsbB.
 KM Oxidoreductase: Redox-active center: Electron transport: Chaperone;
 KM Transmembrane; Inner membrane.
 FT DOMAIN 3 16 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 17 33 POTENTIAL.
 FT DOMAIN 34 51 PERIPLASMIC (POTENTIAL).
 FT TRANSMEM 52 67 POTENTIAL.
 FT DOMAIN 68 74 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 75 92 POTENTIAL.
 FT DOMAIN 93 147 PERIPLASMIC (POTENTIAL).
 FT TRANSMEM 148 165 POTENTIAL.
 FT DISULFID 43 46 REDOX-ACTIVE (BY SIMILARITY).
 FT DISULFID 107 133 REDOX-ACTIVE (BY SIMILARITY).
 FT NON_TER 165 165
 SQ SEQUENCE 165 AA: 18745 MW: 43596DF5A3272D CRC64;
 Query Match 96.7%: Score 29; DB 1; Length 165;
 Best Local Similarity 80.0%: Pred. No. 37;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 1 FPKWR 5
 DB 113 FPSMR 117
 ID E2AA_ECOLI STANDARD: PRT: 259 AA.
 AC P13810;
 DT 01-JAN-1990 (Rel. 13; Created)
 DT 01-JAN-1990 (Rel. 13; Last sequence update)
 DT 01-OCT-1996 (Rel. 34; Last annotation update)
 DE HEAT-LABILE ENTEROTOXIN IIA, A CHAIN PRECURSOR (LT-IIA).
 OS Escherichia coli.
 OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88032841; PubMed=2822667;
 RA Plickett C.L., Weinstein D.L., Holmes R.K.;
 RT "Genetics of type IIA heat-labile enterotoxin of Escherichia coli:
 RT operon fusions, nucleotide sequence, and hybridization studies.";
 RL J. Bacteriol. 169:5180-5187(1987).
 DR

CC -1- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY
 CC THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.
 CC -1- SUBUNIT: HETEROHXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M17894; AAA24093.1; -
 DR PIR: A29831; A29831.
 DR HSSP: P43528; IYIL.
 DR InterPro: IPR001144; Enterotoxin_A.
 DR Pfam: PF01375; Enterotoxin_A; I.
 LR PRINTS: PR00771; ENTEROTOXINA.
 KM Enterotoxin; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 259 HEAT-LABILE ENTEROTOXIN IIA, A CHAIN.
 FT DISULFID 203 215 BY SIMILARITY.
 FT ACT_SITE 128 128
 SQ SEQUENCE 259 AA: 996F311A32CABFA CRC64;
 Query Match 96.7%: Score 29; DB 1; Length 259;
 Best Local Similarity 80.0%: Pred. No. 56;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 1 FPKWR 5
 DB 187 FPSMR 191
 ID GUNY_ERMCH STANDARD: PRT: 332 AA.
 AC P27032;
 DT 01-AUG-1992 (Rel. 23; Created)
 DT 01-AUG-1992 (Rel. 23; Last sequence update)
 DT 01-NOV-1997 (Rel. 35; Last annotation update)
 DE MINOR ENDOGLUCANASE Y PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE
 DE Y) (CELLULOSE Y) (EGY).
 OS Erwinia chrysanthemi.
 OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
 OC Pectobacterium.
 OX NCBI_TaxID=556;
 RN (1)
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-33.
 RC STRAIN-3937;
 RX MEDLINE=92039050; PubMed=1937031;
 RA Giuseppe A., Aymeric J.-L., Caml B., Barras F., Creuzet N.;
 RT "Sequence analysis of the cellulase-encoding cely gene of Erwinia
 RT chrysanthemi: a possible case of interspecies gene transfer.";
 RL Gene 106:109-114(1991).
 CC -1- FUNCTION: REPRESENTS ONLY 3% OF THE GLOBAL CELLULOSE ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY D (FAMILY 8 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M74044; AAA24818.1; -
 DR PIR: J0585; J0585.
 DR

DR InterPro: IPR02037; Glyco_hydro-8.
 DR Pfam: PF01270; Glyco_hydro-8; 1.
 DR PRINTS: PR00735; GLYDRLASE.
 DR PROSITE: PS00812; GLYCOSYL_HYDROL_F8; 1.
 KM Cellulose degradation; Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 23
 FT CHAIN 24 332
 FT ACT_SITE 53 53
 FT ACT_SITE 110 110
 SEQUENCE 332 AA: 37592 MW: 6EA760A2B8227079 CRC64;

Query Match: 96.7%; Score 29; DB 1; Length 332;
 Best Local Similarity 80.0%; Pred. No. 71;
 Matches 4: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
 DB 178 FPXMR 182

RESULT 5
 ID GUN_CELUD STANDARD: PRT: 359 AA.
 AC P18336;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
 DE (CELLULOSE).
 OS Cellulomonas uda.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Micrococcales; Cellulomonadaceae; Cellulomonas.
 NX NCBI_TaxID=1714;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-37.
 RC STRAIN=CB4;
 RA Nakamura K., Misawa N., Kitamura K.;
 RT "Sequence of a cellulase gene of Cellulomonas uda CB4.";
 RL J. Biotechnol. 4:247-254(1986).

CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
 CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
 CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
 CC (2) EXOCELLULOBIODHOLASES THAT CUT THE DISSACCHARIDE CELLOBIOSE
 CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
 CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLOBIOSE AND OTHER
 CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY D (FAMILY 8 OF GLYCOSYL
 CC HYDROLASES).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: M36503; AAA23090.1;
 DR InterPro: IPR02037; Glyco_hydro-8.
 DR Pfam: PF01270; Glyco_hydro-8; 1.
 DR PROSITE: PS00812; GLYCOSYL_HYDROL_F8; 1.
 KM Cellulose degradation; Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 23
 FT CHAIN 24 359
 FT ACT_SITE 53 53
 FT ACT_SITE 110 110
 SEQUENCE 359 AA: 40690 MW: 0445D751B683148 CRC64;

Query Match 96.7%; Score 29; DB 1; Length 359;

Best Local Similarity 80.0%; Pred. No. 76;
 Matches 4: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
 DB 178 FPXMR 182

RESULT 6
 ID CCR5_HUMAN STANDARD: PRT: 372 AA.
 AC P32302; Q14811;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE C-X-C CHEMOKINE RECEPTOR TYPE 5 (CXCR-5) (BURKITT'S LYMPHOMA
 DE RECEPTOR 1) (MONOCYTE-DERIVED RECEPTOR 15) (MDR15).
 GN BLR1 OR CXCR5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG FORM).
 RC TISSUE=Lymphocytes;
 RX MEDLINE=93049615; PubMed=1425907;
 RA Doherty T., Wolf I., Emrich T., Lipp M.;
 RT "Differentiation-specific expression of a novel G protein-coupled
 RT receptor from Burkitt's lymphoma.";
 RL Eur. J. Immunol. 22:2795-2799(1992).
 RN [2]
 RP SEQUENCE FROM N.A. (SHORT FORM).
 RC TISSUE=Blood;
 RX MEDLINE=95366551; PubMed=7639692;
 RA Barella L., Loetscher M., Tobler A., Baggiolini M., Moser B.;
 RT "Sequence variation of a novel heptahelical leucocyte receptor
 RT through alternative transcript formation.";
 RL Biochem. J. 309:773-779(1995).
 RN [3]
 RP LIGAND BINDING.
 RX MEDLINE=98130629; PubMed=9463416;
 RA Legler D.F., Loetscher M., Stuber Roos R., Clark-Lewis I.,
 RA Baggiolini M., Moser B.;
 RT "B cell-attracting chemokine 1, a human CXC chemokine expressed in
 RT lymphoid tissues, selectively attracts B lymphocytes via
 RT BLR1/CXCR5.";
 RL J. Exp. Med. 187:655-660(1998).

CC -1- FUNCTION: CYTOKINE RECEPTOR THAT BINDS TO B1C. BLR1 EXERTS
 CC POSSIBLY A REGULATORY FUNCTION IN BURKITT LYMPHOMA (BL)
 CC LYMPHOMAGENESIS AND/OR B-CELL DIFFERENTIATION. IT IS A POTENTIAL
 CC CANDIDATE FOR CELL-CELL INTERACTION, AND ACTIVATION OF MATURE B-
 CC LYMPHOCYTES IN LYMPHATIC TISSUES.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSION IN MATURE B-CELLS AND BURKITT
 CC LYMPHOMA CELLS.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: X68149; CAA48252.1;
 DR EMBL: X68829; CAA48723.1;
 DR PIR: S26667; S26667.
 DR GCRDB: GCR_0453;
 DR GCRDB: GCR_2072;
 DR GCRDB: GCR_2612;

DR MIM: 601613; -
 DR InterPro: IPR000276; GPCR_Rhodopsn.
 DR Pfam: PF00001; 7tm_1.1.
 DR PRINTS: PR000237; GPCRHHODOPSIN.
 DR PRINTS: PR00564; BURKITSLYMR.
 DR PROSITE: PS00337; G_PROTEIN_RECEPT_FL_1.
 DR PROSITE: PS50262; G_PROTEIN_RECEPT_FL_2.
 DR G-protein coupled receptor; Transmembrane; Glycoprotein; B-cell;
 KW Alternative splicing.
 FT DOMAIN 1 55
 FT TRANSMEM 56 76
 FT DOMAIN 77 88
 FT TRANSMEM 89 109
 FT DOMAIN 110 124
 FT TRANSMEM 125 145
 FT DOMAIN 146 167
 FT TRANSMEM 168 188
 FT DOMAIN 189 219
 FT TRANSMEM 220 240
 FT DOMAIN 241 259
 FT TRANSMEM 260 280
 FT DOMAIN 281 304
 FT TRANSMEM 305 325
 FT DOMAIN 326 372
 FT CARBOHYD 28
 FT CARBOHYD 126
 FT DISULFID 196
 FT DISULFID 202
 FT VARSPFLC 1
 FT CONFLICT 344
 FT CONFLICT 344
 FT SEQUENCE 372 AA: 41955 MW: 60F84C839492ACCF CRC64;

Query Match 96.7%; Score 29; DB 1; Length 372;
 Best Local Similarity 80.0%; Pred. No. 79;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPKWR 5
 1111
 DB 352 FPKWR 356

RESULT 7
 MCP_HUMAN
 ID MCP_HUMAN STANDARD: PRT: 377 AA.
 AC P15529;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE MEMBRANE COFACTOR PROTEIN PRECURSOR (CD46 ANTIGEN) (TROPHOBLAST
 DE LEUCOCYTE COMMON ANTIGEN) (TLX).
 GN MCP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI-TaxID:9606;
 111
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 35-58.
 RX MEDLINE-88286080; PubMed-3360937;
 RA Lublin D.M., Liszewski M.K., Post T.W., Arce M.A., Le Beau M.M.,
 RA Reberlisch M.B., Lemons R.S., Seya T., Atkinson J.P.;
 RT "Molecular cloning and chromosomal localization of human membrane
 RT cofactor protein (MCP). Evidence for inclusion in the multigene
 RT family of complement-regulatory proteins.";
 RL J. Exp. Med. 168:181-194(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE-93119658; PubMed-8418811;
 RA Ceroni F., Fenichel P., Akhondji C., Hsi B.L., Rossi B.;
 RT "Characterization of a cDNA clone coding for human testis membrane
 RT cofactor protein (MCP, CD46).";
 RL Mol. Reprod. Dev. 34:107-113(1993).
 RN [3]

RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RX MEDLINE-91267562; PubMed-2050389;
 RA Purcell D.F., Russell S.M., Deacon N.J., Brown M.A., Hooker D.J.,
 RA McKenzie I.F.;
 RT "Alternatively spliced RNAs encode several isoforms of CD46 (MCP), a
 RT regulator of complement activation.";
 RL Immunogenetics 33:335-344(1991).
 RN [4]
 RP SEQUENCE OF 1-34 FROM N.A.
 RX MEDLINE-94014356; PubMed-7691939;
 RA Cul W., Hourcade D., Post T., Greenland A.C., Atkinson J.P.,
 RA Kumar V.;
 RT "Characterization of the promoter region of the membrane cofactor
 RT protein (CD46) gene of the human complement system and comparison to
 RT a membrane cofactor protein-like genetic element.";
 RL J. Immunol. 151:4137-4146(1993).
 RN [5]
 RP ALTERNATIVE SPLICING.
 RX MEDLINE-92289809; PubMed-1601037;
 RA Russell S.M., Sparrow R.L., McKenzie I.F.C., Purcell D.F.J.;
 PT "Tissue-specific and allelic expression of the complement regulator
 PT CD46 is controlled by alternative splicing.";
 RL Eur. J. Immunol. 22:1513-1518(1992).
 CC -I- FUNCTION: MCP IS AN IMPORTANT MEMBRANE PROTEIN FOR PROTECTING HOST
 CC CELLS FROM DAMAGE BY COMPLEMENT. MCP POSSESSES COFACTOR ACTIVITY
 CC FOR INACTIVATION OF C3B AND C4B BY SERUM FACTOR I. THIS WOULD
 CC ALLOW MCP TO REGULATE THE COMPLEMENT SYSTEM AT THE LEVEL OF THE C3
 CC CONVERTASES. MAY BE INVOLVED IN THE FUSION OF THE SPERMATOZOA WITH
 CC THE OOCYTE DURING FERTILIZATION. MAY ACT BY PREVENTING COMPLEMENT
 CC ACTIVATION AT THE FETO-MATERNAL INTERFACE ON THE
 CC SYNCYTOTROPHOBLAST LAYER OF PLACENTA.
 CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -I- ALTERNATIVE PRODUCTS: AT LEAST 14 FORMS ARE PRODUCED BY
 CC ALTERNATIVE SPLICING. THE FORM SHOWN IS THE TRANSCRIPT A. THE
 CC GAMMA ISOFORM (TRANSCRIPTS A AND B) IS PREFERENTIALLY EXPRESSED IN
 CC EBV-B CELLS AND LEUKEMIC CELLS. THE 66 KDA ALPHA ISOFORM AND THE
 CC 56 KDA BETA ISOFORM FOUND IN ALL TISSUES EXCEPT SPERM CORRESPOND
 CC TO THE TRANSCRIPTS C AND D. AND THE TRANSCRIPTS E AND F
 CC RESPECTIVELY. THE EXON 9 IS SPECIFICALLY DELETED IN SOME
 CC PLACENTAE. THE EPSILON ISOFORM CORRESPONDS TO THE TRANSCRIPTS I
 CC AND J. SPERMATOZOA DELTA ISOFORM (TRANSCRIPT N) DELETE BOTH EXONS
 CC 12 AND 13. ALL TISSUES DIFFERENTIALLY SPLICE EXON 13.
 CC -I- TISSUE SPECIFICITY: EXPRESSED BY ALMOST ALL CELLS EXCEPT
 CC ERYTHROCYTES AND SOME BONE MARROW CELLS.
 CC -I- PTM: EXTENSIVELY O-GLYCOSYLATED IN THE SER/THR-RICH DOMAIN
 CC (PROBABLE).
 CC -I- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
 CC -I- DATABASE: NAME=PROW; NOTE=CD guide CD46 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd46.htm".
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib.ch).
 CC -----
 DR EMBL: Y00651; CAA68675.1; -
 DR EMBL: S51940; AAB24802.1; -
 DR EMBL: M58050; AAB24813.1; -
 DR EMBL: A18585; CAA01400.1; -
 DR EMBL: S65879; AAD13968.1; -
 DR PIR: S01896; S01896.
 DR HSSP: P10998; IIVC.
 DR MIM: 120920; -
 DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF000084; sushi_4.
 DR SMART: SM00032; CCP: 4.
 KM Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
 KW Sushi; Alternative splicing.
 FT SIGNAL 1 34
 FT CHAIN 35 377 MEMBRANE COFACTOR PROTEIN.

```

FT DOMAIN 35 328 EXTRACELLULAR (POTENTIAL).
FT TRANSEM 329 351 POTENTIAL.
FT DOMAIN 352 377 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 355 95 SUSHI 1.
FT DOMAIN 98 158 SUSHI 2.
FT DOMAIN 161 224 SUSHI 3.
FT DOMAIN 227 284 SUSHI 4.
FT DOMAIN 287 311 SER/THR-RICH.
FT DISULFID 35 80 BY SIMILARITY.
FT DISULFID 64 94 BY SIMILARITY.
FT DISULFID 99 141 BY SIMILARITY.
FT DISULFID 127 157 BY SIMILARITY.
FT DISULFID 162 210 BY SIMILARITY.
FT DISULFID 191 223 BY SIMILARITY.
FT DISULFID 228 270 BY SIMILARITY.
FT CARBOHYD 256 283 BY SIMILARITY.
FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 273 273 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPPLIC 286 300 MISSING (IN A SECOND FORM).
FT VARSPPLIC 340 352 MISSING (IN ISOFORM M).
FT VARSPPLIC 353 361 YLORRKKKG -> DIFKGRCKCKVVELNPLRLNQLQ
OSREAE (IN ISOFORM M).
FT VARSPPLIC 362 377 TYLDTETREVEKFTSL -> KADGCAEVATYQKSTTPAEQ
RG (IN ISOFORM B/D/E/H/J/L).
FT VARSPPLIC 362 377 MISSING (IN ISOFORM M AND ISOFORM N).
FT VARSPPLIC 340 361 MISSING (IN ISOFORM N).
FT VARSPPLIC 339 339 I -> ICKQVVELNPLRLNQLQOSREAE (IN
ISOFORM N).
SQ SEQUENCE 377 AA: 42247 MW: 2CAGF61752570B57 CRC64:

Query Match 96.7%: Score 29: DB 1: Length 377:
Best Local Similarity 80.0%: Pred. No. 80:
Matches 4: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

Qy 1 FPXWR 5
Db 11 FPSMR 15

RESULT 8
2229_HUMAN STANDARD: PRT: 420 AA.
ID 2229_HUMAN
AC 09JUN7:
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ZINC FINGER PROTEIN 229 (FRAGMENT).
GN ZNF229.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RA Shannon M., Branscomb E., Hauser L., Gordon L., Ashworth L.,
Stubs L.;
RT Differential expansion of homologous zinc-finger gene families in
human chromosome 19q13.2 and mouse chromosome 7.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN 12
RP SEQUENCE FROM N.A.
RA Kodyliani V., Ge Y., Krummel G.K., Kvistad E., Grable J.,
Severin J., Gordon L., Shannon M., Brower A., Olsen A.S., Smith L.M.;
RT Sequence analysis of a 1mb region in 19q13.2 containing a zinc finger
gene cluster.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL: AF192979; AAF07964.1; -
DR EMBL: AC084239; AAG23970.1; -
DR InterPro: IPR001909; KRAB.
DR InterPro: IPR000822; Znf-C2H2.
DR Pfam: PF01352; KRAB.1.
DR Pfam: PR00096; zf-C2H2.2.
DR PRINTS: PR00046; ZINC_FINGER.
DR SMART: SM00349; KRAB.1.
DR SMART: SM00355; Znf_C2H2.2.
DR PROSITE: PS00805; KRAB.1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 2.
DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 3.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 34 108 KRAB.
FT DOMAIN 349 420 ZINC_FINGERS.
FT ZN_FING 349 371 C2H2-TYPE.
FT ZN_FING 377 399 C2H2-TYPE.
FT ZN_FING 405 420 C2H2-TYPE.
FT NON_TER 420 420
SQ SEQUENCE 420 AA: 48022 MW: FFA138BA44A2A14 CRC64:

Query Match 96.7%: Score 29: DB 1: Length 420:
Best Local Similarity 80.0%: Pred. No. 89:
Matches 4: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

Qy 1 FPXWR 5
Db 173 FPAMR 177

RESULT 9
HS16_CAEEL STANDARD: PRT: 143 AA.
ID HS16_CAEEL
AC P06581; P02514:
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE HEAT SHOCK PROTEIN HSP16-41.
GN HSP16-41.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pseudocercariae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN 11
RP SEQUENCE FROM N.A.
RA Jones D., Russnak R.H., Kay R.J., Candido E.P.M.;
RT "Structure, expression, and evolution of a heat shock gene locus in
Caenorhabditis elegans that is flanked by repetitive elements.";
RL J. Biol. Chem. 261:12006-12015(1986).
RN 12
RP SEQUENCE OF 47-143 FROM N.A.
RA MEDLINE-83220736; PubMed-6190129;
RX Russnak R.H., Jones D., Candido E.P.M.;
RT Cloning and analysis of cDNA sequences coding for two 16 kilodalton
heat shock proteins (hsps) in Caenorhabditis elegans: homology with
the small hsps of Drosophila.;
RL Nucleic Acids Res. 11:3187-3205(1983).
CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC
 CC EMBL: M14334; AAA28070.1; ALT-SEQ.
 CC EMBL: X01577; CAA25732.1; -
 CC PIR: A38884; HHKMA1.
 CC PIR: A25199; A25199.
 CC InterPro: IPR002068; Crystallin_HSP20.
 CC Pfam: PF00001; HSP20.1.
 CC PROSITE: PS01031; HSP20.1.
 CC Heat shock; Multigene family.
 CC SEQUENCE 143 AA; 16252 MW; C1D0F59D26E36C24 CRC64;

Query Match 93.3%; Score 28; DB 1; Length 143;
 Best Local Similarity 80.0%; Pred. No. 50;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
 11 11
 27 FPYWR 31

Db

RESULT 10
 ID HS17 CAEEL STANDARD: PRT: 143 AA.
 AC P02513;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE HEAT SHOCK PROTEIN HSP16-48.
 GN HSP16-48.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 11 11
 RN SEQUENCE FROM N.A.
 RX MEDLINE=85295957; PubMed=4033652;
 RA Russnak R.H.; Candido E.P.M.;
 RT "Locus encoding a family of small heat shock genes in Caenorhabditis
 RT elegans: two genes duplicated to form a 3.8-kilobase inverted
 RT repeat";
 RL Mol. Cell. Biol. 5:1268-1278(1985).
 RN 121
 RP SEQUENCE OF 9-143 FROM N.A.
 RX MEDLINE=83220736; PubMed=6190129;
 RA Russnak R.H.; Jones D.; Candido E.P.M.;
 RT "Cloning and analysis of cDNA sequences coding for two 16 kilodalton
 RT heat shock proteins (hsps) in Caenorhabditis elegans: homology with
 RT the small hsps of Drosophila";
 RL Nucleic Acids Res. 11:3187-3205(1983).
 CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
 CC FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC
 CC EMBL: X01577; AAA28069.1; -
 CC EMBL: X01576; CAA25731.1; -
 CC EMBL: X01863; AAA28064.1; -
 CC PIR: A02916; HHKMA8.
 CC PIR: A24289; A24289.
 CC InterPro: IPR002068; Crystallin_HSP20.

DR Pfam: PF00011; HSP20.1.
 DR PROSITE: PS01031; HSP20.1.
 DR Heat shock; Multigene family.
 DR SEQUENCE 143 AA; 16299 MW; 0D5596DFEE5B3318 CRC64;

Query Match 93.3%; Score 28; DB 1; Length 143;
 Best Local Similarity 80.0%; Pred. No. 50;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
 11 11
 27 FPYWR 31

Db

RESULT 11
 ID HNL MANES STANDARD: PRT: 257 AA.
 AC P52705;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE (S)-ACETONE-CYANOHYDRIN LYASE (EC 4.1.2.39) ((S)-HYDROXYNITRILE LYASE)
 DE ((S)-HYDROXYNITRILASE) (OXNITRILASE).
 GN HNL.
 OS Manihot esculenta (Cassava) (Manihot).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Malpighiales; Euphorbiaceae; Manihot.
 OX NCBI_TaxID=3983;
 11 11
 RN SEQUENCE FROM N.A., AND SEQUENCE OF 1-36 AND 168-191.
 RP TISSUE=Cotyledon;
 FC MEDLINE=94263231; PubMed=8203915;
 KA Hughes J.; Carvalho F.J.P.D.C.; Hughes M.A.;
 KL Arch. Biochem. Biophys. 311:496-502(1994).
 RT "Purification, characterization, and cloning of alpha-hydroxynitrile
 RT lyase from cassava (Manihot esculenta Crantz).";
 CC Arch. Biochem. Biophys. 311:496-502(1994).
 CC -1- FUNCTION: INVOLVED IN CYANOGENESIS. THE RELEASE OF HCN FROM
 CC INJURED TISSUES. DECOMPOSES A VARIETIES OF (R) OR (S) CYANOHYDRINS
 CC INTO HCN AND THE CORRESPONDING ALDEHYDES AND KETONES. THE NATURAL
 CC SUBSTRATE OF THIS ENZYME IS (S)-ACETONE CYANOHYDRIN.
 CC -1- CATALYTIC ACTIVITY: (S)-2-HYDROXYISOBUTYRONITRILE - CYANIDE +
 CC ACETONE.
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- SIMILARITY: DISTANT RELATIONSHIP WITH TYPE-B
 CC CARBOXYLESTERASE/LIPASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC
 CC EMBL: Z29091; CAA82334.1; -
 CC HSP: P52704; LYAS.
 DR InterPro: IPR000073; Abhydrolase.
 DR InterPro: IPR000379; Est_Lip_thioest_actsite.
 DR Pfam: PF00561; abhydrolase.1.
 KW Lyase.
 FT INIT_MEN 0 0
 FT ACT_SITE 79 79 BY SIMILARITY.
 FT ACT_SITE 207 207 BY SIMILARITY.
 FT ACT_SITE 235 235 BY SIMILARITY.
 FT ACT_SITE 235 235 BY SIMILARITY.
 SQ SEQUENCE 257 AA; 29240 MW; 98B3E160ACB338C5 CRC64;

Query Match 93.3%; Score 28; DB 1; Length 257;
 Best Local Similarity 80.0%; Pred. No. 87;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X76901; CAA54223.1; -
 DR PIR: B36971; B36971.
 DR PIR: S40129; S40129.
 DR InterPro: IPR003187; PLA1.
 DR Pfam: PF02253; PLA1; 1.
 KM Hydrolyase; Lipid degradation: Outer membrane: Signal; Calcium.
 FT SIGNAL 1 20 BY SIMILARITY.
 FT CHAIN 21 286 PHOSPHOLIPASE A1.
 FT ACT_SITE 161 161 BY SIMILARITY.
 SO SEQUENCE 286 AA; 32544 MW; 3E39F863085108A3 CRC64;

 Query Match 93.3%; Score 28; DB 1; Length 286;
 Best Local Similarity 80.0%; Pred. No. 96;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

 OY 1 FPXWR 5
 Db 92 FPLWR 96

 RESULT 15
 PAL_ECOLI STANDARD: PRT: 289 AA.
 AC P00631;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PHOSPHOLIPASE A1 PRECURSOR (EC 3.1.1.32) (DEREGENT-RESISTANT
 DE PHOSPHOLIPASE A) (DR-PHOSPHOLIPASE A) (PHOSPHATIDYLCHOLINE 1-
 DE ACYLHYDROLASE) (OUTER MEMBRANE PHOSPHOLIPASE A) (OM PLA).
 GN PLDA OR B3821 OR Z5342 OR ECS4751.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae;
 OC Escherichia
 CC NCBI_TaxID=562, 83334;
 OX 11
 RP SEQUENCE FROM N.A.
 RA MEDLINE=85157492; PubMed=6397464;
 RA Homma H., Kobayashi T., Chiba N., Karasawa K., Mizushima H., Kudo T.,
 RA Inoue K., Ikeda H., Sekiguchi M., Nojima S.;
 RT "The DNA sequence encoding plda gene, the structural gene for
 RT detergent-resistant phospholipase A of E. coli.";
 RL J. Biochem. 96:1655-1664(1984).
 RN [12]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=92358234; PubMed=1379743;
 RA Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;
 RT "Analysis of the Escherichia coli genome: DNA sequence of the region
 RT from 84.5 to 86.5 minutes.";
 RL Science 257:771-778(1992).
 RN [13]
 RP REVISION TO 14-15;
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [14]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yan G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533(2001).
 RN [15]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 RN [16]
 RP SEQUENCE OF 30-289 FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=85003590; PubMed=6383820;
 RA de Gues P., Verheij H.M., Riegman N.H., Hoekstra W.P.M., de Haas G.H.;
 RT "The pro- and mature forms of the E. coli K-12 outer membrane
 RT phospholipase A are identical.";
 RL EMBO J. 3:1799-1802(1984).
 RN [17]
 RP SEQUENCE OF 174-289 FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=87115164; PubMed=3027506;
 RA Irino N., Nakayama K., Nakayama H.;
 RT "The recQ gene of Escherichia coli K12: primary structure and
 RT evidence for SOS regulation.";
 RL Mol. Gen. Genet. 205:298-304(1986).
 RN [18]
 RP MUTAGENESIS OF SER-172.
 RC MEDLINE=94131966; PubMed=8300539;
 RA Brok R.G.P.M., Brinkman E., van Boxtel R., Bekkers A.C.A.P.,
 RA Verheij H.M., Tomassen J.;
 RT "Molecular characterization of enterobacterial plda genes encoding
 RT outer membrane phospholipase A.";
 RL J. Bacteriol. 176:861-870(1994).
 RN [19]
 RP ACTIVE SITE SER-164, AND PARTIAL SEQUENCE.
 RC MEDLINE=91249806; PubMed=2040286;
 RA Horrevorts A.D.G., Verheij H.M., de Haas G.H.;
 RT "Inactivation of Escherichia coli outer-membrane phospholipase A by
 RT the affinity label hexadecanesulfonyl fluoride. Evidence for an
 RT active-site serine.";
 RL Eur. J. Biochem. 198:247-253(1991).
 CC -I- FUNCTION: HYDROLYSIS OF PHOSPHATIDYLCHOLINE WITH PHOSPHOLIPASE
 CC A2 (EC 3.1.1.4) AND PHOSPHOLIPASE A1 (EC 3.1.1.32) ACTIVITIES.
 CC REQUIRED FOR EFFICIENT SECRETION OF BACTERIOCINS, SEEMS TO BE
 CC DORMANT IN NORMAL GROWING CELLS.
 CC -I- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 1-ACYLGLYCERO-
 CC PHOSPHOCHOLINE + A FATTY ACID ANION.
 CC -I- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 2-ACYLGLYCERO-
 CC PHOSPHOCHOLINE + A FATTY ACID ANION.
 CC -I- COFACTOR: REQUIRES CALCIUM IONS FOR ACTIVITY.
 CC -I- SUBCELLULAR LOCATION: OUTER MEMBRANE; ONE OF THE VERY FEW ENZYMES
 CC LOCATED THERE.
 CC -I- INDUCTION: BY MEMBRANE DAMAGING, FOR EXAMPLE, BY PHAGE-INDUCED
 CC LYSIS OR TEMPERATURE SHOCK.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: X02143; CAA26081.1; -
DR EMBL: M87049; AAA67617.1; -
DR EMBL: AE000458; AAC76824.1; -
DR EMBL: AE005613; AAC59017.1; -
DR EMBL: AP002567; BAB38174.1; -
DR EMBL: M30198; AAA24516.1; -
DR PIR: A00771; PSECA.
DR PIR: A2133; PSECA1.
DR PIR: S30711; S30711.
DR Ecogene: EG10738; pIDA.
DR InterPro: IPR003187; PLA1.
DR Pfam: PF02253; PLA1; 1.
KW Hydrolase; lipid degradation; Outer membrane; Signal; Calcium;
Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 289 PHOSPHOLIPASE A1.
FT ACT_SITE 164 164
FT MUTAGEN 172 172 S->F: INACTIVE PROTEIN.
FT CONFLICT 14 15 LP -> FA (IN REF. 2).
FT CONFLICT 30 33 DAPA -> MTRQ (IN REF. 6).
SQ SEQUENCE 289 AA: 33163 MW: A688AD32AA60F218 CRC64;

Query Match 93.3%; Score 28; DB 1; Length 289;
Best Local Similarity 80.0%; Pred. No. 97;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
|||
DB 95 PPLMR 99

Search completed: February 27, 2002, 11:42:48
Job time: 547 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:50:14 ; Search time 281.76 Seconds
(without alignments)
2.596 Million cell updates/sec

Title: US-09-446-109a-14
Perfect score: 30
Sequence: 1 FPXMR 5

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database:

1: SP archaea: *
2: SP bacteria: *
3: SP fungi: *
4: SP human: *
5: SP invertebrate: *
6: SP mammal: *
7: SP_mhc: *
8: SP organelle: *
9: SP phage: *
10: SP plant: *
11: SP rodent: *
12: SP virus: *
13: SP vertebrate: *
14: SP unclassified: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
1	29	96.7	101	5	09V671	09V671 drosophila
2	29	96.7	113	5	039293	039293 brassica na
3	29	96.7	157	5	09X75	09X75 caenorhabd
4	29	96.7	158	4	09NMW3	09NMW3 homo sapien
5	29	96.7	158	4	09NMW2	09NMW2 homo sapien
6	29	96.7	229	8	037633	037633 rhagoletis
7	29	96.7	245	2	09PDE1	09PDE1 xytelella fas
8	29	96.7	273	10	09SUV9	09SUV9 arbidopsi
9	29	96.7	292	2	09H873	09H873 pseudomonas
10	29	96.7	321	11	09PDE7	09PDE7 mus muscul
11	29	96.7	332	2	09APD5	09APD5 erwinia chr
12	29	96.7	333	2	09L309	09L309 erwinia rha
13	29	96.7	334	10	09FTM4	09FTM4 oryza sativ
14	29	96.7	349	4	015429	015429 homo sapien
15	29	96.7	417	10	09SNA1	09SNA1 arabidopsi
16	29	96.7	466	2	09FDR4	09FDR4 zymomonas m
17	29	96.7	497	2	086742	086742 streptomyces
18	29	96.7	497	2	091702	091702 pseudomonas
19	29	96.7	501	5	016923	016923 caenorhabd

20	29	96.7	502	5	09G015	09G015 caenorhabd
21	29	96.7	555	1	029748	029748 archaeoglob
22	29	96.7	581	4	09BR01	09BR01 homo sapien
23	29	96.7	612	4	09H8Y2	09H8Y2 homo sapien
24	29	96.7	621	4	09H9Y1	09H9Y1 homo sapien
25	29	96.7	625	2	066676	066676 aquilex aeo
26	29	96.7	634	4	09NM19	09NM19 homo sapien
27	29	96.7	642	4	09NMX5	09NMX5 homo sapien
28	29	96.7	642	4	025396	025396 helicobacte
29	29	96.7	642	2	09ZLF3	09ZLF3 helicobacte
30	29	96.7	650	5	017866	017866 caenorhabd
31	29	96.7	735	2	09H063	09H063 pseudomonas
32	29	96.7	778	4	09U9K6	09U9K6 caenorhabd
33	29	96.7	811	4	09BWX2	09BWX2 homo sapien
34	29	96.7	1071	11	09D2K4	09D2K4 mus muscul
35	29	96.7	1108	13	09PW00	09PW00 tetraodon f
36	29	96.7	1117	5	09U9K7	09U9K7 caenorhabd
37	29	96.7	1451	2	045336	045336 bordetella
38	29	96.7	1451	2	045044	045044 bordetella
39	29	96.7	1464	2	09S5D5	09S5D5 bordetella
40	29	93.3	68	2	P73858	P73858 synchocyst
41	28	93.3	84	2	P74471	P74471 synchocyst
42	28	93.3	84	2	055568	055568 synchocyst
43	28	93.3	93	2	P73028	P73028 synchocyst
44	28	93.3	103	10	09M045	09M045 arabidopsi
45	28	93.3	110	6	09G1J9	09G1J9 canis fami

ALIGNMENTS

RESULT 1
ID 09V671 PRELIMINARY: PRT: 101 AA.
AC 09V671: 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CG13174 PROTEIN.
CN CG13174.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:
OC Pterygota: Neoptera: Endopterygota: Diptera: Brachycera: Muscomorpha:
OC Ephydroidea: Drosophilidae: Drosophila.
UX NCBI_TaxID=7227;
FN [1]
RP SEQUENCE FROM N.A.
KC STRAIN=BERKELEY;
KC MEDLINE=20196006; Pubmed=10731132;
KA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
KA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Calle R.F.,
KA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
KA Sulten G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
KA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
KA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
KA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
KA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
KA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolintsov S.,
KA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
KA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
KA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
KA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
KA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
KA Durbin K.J., Evangelista C.C., Ferraz C., Fertile W.S., Fleischmann W.,
KA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
KA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
KA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
KA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
KA Jatala M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
KA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
KA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
KA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
KA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Piltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Palzer K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spletter E., Spralling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT The genome-sequence of *Drosophila melanogaster*.
RL Science 287:2185-2195(2000).
DR EMBL: AE003823; AAF58560.1;
DR FlyBase: FBgn0033694; CG13174.
SQ SEQUENCE 101 AA: 11820 MW: D5C4D8B514CB89B CRC64;

Query Match
Best Local Similarity 96.7%; Score 29; DB 5; Length 101;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPKMR 5
DB 75 FPKMR 79

RESULT 2
O39293 PRELIMINARY: PRT: 113 AA.
AC O39293;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE ELONGATION FACTOR EF-1A (FRAGMENT).
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3708;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN-SAMOURAI; TISSUE=ETIOLATED SEEDLINGS, (5-DAYS-OLD);
RA Saez-Vasquez J.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO GTP-BINDING ELONGATION FACTOR FAMILY.
DR EMBL: U21744; AAA6366.1;
DR Mendel: 15767; Brana:1139;15767.
DR InterPro: IPR000795; GTP_EFTU.
DR Pfam: PF00009; GTP_EFTU; 1.
KM Elongation factor: GTP-binding; Protein biosynthesis.
FT NON_TER 1
SQ SEQUENCE 113 AA: 12560 MW: 1497830707223948 CRC64;

Query Match
Best Local Similarity 96.7%; Score 29; DB 10; Length 113;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPKMR 5
DB 103 FPKMR 107

RESULT 3
O9XTY5 PRELIMINARY: PRT: 157 AA.
AC O9XTY5;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE T03D8.2 PROTEIN.

GN T03D8.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pterodermidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN 111
RP SEQUENCE FROM N.A.
RA Mortimore B.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN 121
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906396;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rikken L., Roopra A., Saunders D., Showkeen R.,
RA Smailson N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterson R.,
RA Watson A., Weinstein L., Wilkinson-Sproat J., Wolfdan P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RL Nature 368:32-38(1994).
DR EMBL: Z92838; CAB07406.1;
DR InterPro: IPR000230; Ribosomal_S12.
DR Pfam: PF00164; Ribosomal_S12; 1.
DR PRINTS: PR01034; Ribosomal_S12.
DR ProDom: PD000576; Ribosomal_S12; 1.
SQ SEQUENCE 157 AA: 17370 MW: 6C6A837282EB26C CRC64;

Query Match
Best Local Similarity 96.7%; Score 29; DB 5; Length 157;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPKMR 5
DB 40 FPKMR 44

RESULT 4
O9NNM3 PRELIMINARY: PRT: 158 AA.
AC O9NNM3;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE MEMBRANE COPOLYMER PROTEIN CD46 VARIANT (FRAGMENT).
GN MCP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 111
RP SEQUENCE FROM N.A.
RX MEDLINE=20227770; PubMed=10751138;
RA Kusuhara K., Sasaki Y., Nakao F., Ihara K., Hattori H., Yamashita S.,
RA Nishii K., Koido N., Aiba H., Takeshita K., Hara T.,
RT "Analysis of measles virus binding sites of the CD46 gene in patients
PT with subacute sclerosing panencephalitis.";
RL J. Infect. Dis. 181:1447-1449(2000).
DR EMBL: AF209713; AAF73845.1;
DR InterPro: IPR000436; Sush1_SCR_CCP.
DR Pfam: PF00084; sush1; 2.
DR SMART: SM00032; CCP; 2.
FT VARIANT 59
FT NON_TER 59
SQ SEQUENCE 158 AA: 18098 MW: 4E3F07EBC5C454F5 CRC64;

Query Match
Best Local Similarity 96.7%; Score 29; DB 4; Length 158;

Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FPXWR 5
11 11
DB 11 FPSWR 15

RESULT 5

O9NNW2 PRELIMINARY; PRT: 158 AA.

AC O9NNW2: 01-OCT-2000 (TREMURel. 15, Created)
DT 01-OCT-2000 (TREMURel. 15, Last sequence update)
DE 01-JUN-2001 (TREMURel. 17, Last annotation update)
GN MEMBRANE COFACTOR PROTEIN CD46 VARIANT (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2022770; PubMed=10751138;
RA Kusunara K., Sasaki Y., Nakao F., Ihara K., Hattori H., Yamashita S.,
RA Nihel K., Koide N., Alba H., Takeshita K., Hara T.;
RT "Analysis of measles virus binding sites of the CD46 gene in patients
RT with subacute sclerosing panencephalitis."
RL J. Infect. Dis. 181:1447-1449(2000).
DR EMBL: AF209714; AAF73846.1; -;
DR InterPro: IPR000436; Sushl_SCR_CCP.
DR Pfam: PF00084; sushl: 2.
DR SMART: SM00032; CCP: 2.
FT NON_TER 158
FT SEQUENCE 158 AA; 18126 MW; E2CF38EC3A16A9DA CRC64;

Query Match 96.7%; Score 29; DB 4; Length 158;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FPXWR 5
11 11
DB 11 FPSWR 15

RESULT 6

O37633 PRELIMINARY; PRT: 229 AA.

AC O37633: 01-NOV-1996 (TREMURel. 01, Created)
DT 01-NOV-1996 (TREMURel. 01, Last sequence update)
DE 01-JUN-2001 (TREMURel. 17, Last annotation update)
GN CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1) (FRAGMENT).
OS Cytochrome c.
OC Rhagolepis boylei.
OS Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Tephritidae; Tephritidae; Rhagolepis.
OX NCBI_Taxid=43419;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97159559; PubMed=9007018;
RA Smith J.J., Bush G.L.;
RT "Phylogeny of the genus Rhagolepis (Diptera: Tephritidae) inferred
RT from DNA sequences of mitochondrial cytochrome oxidase II."
RL Mol. Phylogenet. Evol. 7:33-43(1997).
GN FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
CC A CENTER TO THE BINUCLEAR CENTER OF THE CATALYTIC SUBUNIT 1 (BY

CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O + 4
CC FERROCYTOCHROME C.
CC -1- COFACTOR: COPPER A AND HEME GROUP.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE (BY SIMILARITY).
CC -1- SIMILARITY: TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
DR EMBL: U53254; AAB50355.1; -;
DR HSSP: P08306; 1AFL
DR InterPro: IPR001505; COX2.
DR InterPro: IPR002429; CYL_C-ox_2.
DR Pfam: PF00116; COX2; 1.
DR PRINTS: PR01166; CYCOXIDASEIL.
DR ProDom: PD000131; COX2; 1.
DR PROSITE: PS00078; COX2; 1.
KW Copper; Inner membrane; Mitochondrion; Oxidoreductase;
KW Respiratory chain; Transmembrane.
FT METAL 159
FT METAL 163
FT METAL 167
FT METAL 207
FT METAL 207
FT NON_TER 229
FT SEQUENCE 229 AA; 26389 MW; 912D1B2736330698 CRC64;

OY 1 FPXWR 5
11 11
DB 78 FPSWR 82

RESULT 7

O9PDE1 PRELIMINARY; PRT: 245 AA.

AC O9PDE1: 01-OCT-2000 (TREMURel. 15, Created)
DT 01-OCT-2000 (TREMURel. 15, Last sequence update)
DE 01-MAR-2001 (TREMURel. 16, Last annotation update)
GN HYPOTHETICAL PROTEIN XF1438.
GN XF1438.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_Taxid=2371;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Britton R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,
RA Coutinho N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Mocena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohnel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Lairol F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Marques M.A., Madeira A.M.B.N., Madeira H.M.F., Martino C.L.,
RA Menck C.F.M., Miracca E.C., Miyaki C.V., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A.J., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Paixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,

RA da Silveira J.F., Silvestri M.L.Z., Silveira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsunako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovsky-Almeida S., Vettore A.L.,
 RA Zago M.A., Zaitz M., Meidanis J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*.";
 RL Nature 406:151-159(2000).
 DR EMBL: AE003974; AAF84247.1; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 245 AA; 27193 MW; 16C8ADB9FC490455 CRC64;

Query Match: 96.7%; Score 29; DB 2; Length 245;
 Best Local Similarity 80.0%; Pred. No. 2.8e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPKWR 5
 DB 197 FPKWR 201

RESULT 8
 O9SUV9 PRELIMINARY: PRT: 273 AA.
 AC O9SUV9:
 DT 01-MAY-2000 (TEMBLrel. 13, Created)
 DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
 DT 01-OCT-2000 (TEMBLrel. 15, Last annotation update)
 DE HYPOTHEICAL 31.2 KDA PROTEIN.
 CN FPK2.180 OR ATAG22600.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702.
 RN 11
 RP SEQUENCE FROM N.A.
 RA Bevan M., Wedler H., Wambutt R., Bancroft I., Mewes H.W.,
 RA Meyer K.F.X., Lemcke K., Schueller C.;
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RN 12
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 RN 13
 RP SEQUENCE FROM N.A.
 RA Wedler H., Wambutt R., Mewes H.W., Lemcke K., Meyer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN 14
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL033545; CA22165.1; -;
 DR EMBL: AL01557; CAB79215.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 273 AA; 31248 MW; 7E644F686CE8EBB0 CRC64;

Query Match: 96.7%; Score 29; DB 10; Length 273;
 Best Local Similarity 80.0%; Pred. No. 3.1e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPKWR 5
 DB 79 FPKWR 83

RESULT 9
 O9HTB3 PRELIMINARY: PRT: 292 AA.
 AC O9HTB3:
 DT 01-MAR-2001 (TEMBLrel. 16, Created)
 DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)

DE HYPOTHETICAL PROTEIN PA5488.
 GN PA5488.
 OS Pseudomonas aeruginosa.
 CC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 CC Pseudomonas.
 OX NCBI_TaxID=287;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
 RA Gardner R.L., Collier L., Tolentino K.E., Westbrock-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Mong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 DR EMBL: AE004961; AAG08873.1; -;
 DR Interpro: IPR000504; RRM.
 DR PROSITE: PS00030; RRM_RNP_1; UNKNOWN_1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 292 AA; 32949 MW; 14610337E8CCDFD4 CRC64;

Query Match: 96.7%; Score 29; DB 2; Length 292;
 Best Local Similarity 80.0%; Pred. No. 3.3e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPKWR 5
 DB 231 FPKWR 235

RESULT 10
 O9D6L7 PRELIMINARY: PRT: 321 AA.
 AC O9D6L7:
 DT 01-JUN-2001 (TEMBLrel. 17, Created)
 DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
 DE 2310076014RIK PROTEIN.
 GN 2310076014RIK.
 OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE= Tongue;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawal J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Pleischmann W., Gaasterland T., Glisi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsch G.,
 RA Blake J., Boffelli D., Bojunga N., Carinici P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazarelli J., Mombauts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Welter C., Whitaker C., Wilmberg L.,
 RA Wyshaw-Borts A., Yoshida K., Hasegawa Y., Kawaji H., Kohisaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL: AK010208; BAB26769.1; -;

DR MGD:1919189; 2310076014rik.
 DR Interpro: IPR000847; HTH_LYSR.
 DR PROSITE: PS00044; HTH_LYSR.FAMILY: UNKNOWN_1.
 SO SEQUENCE 321 AA; 35038 MW; 6E27EBC746552DFB CRC64;

Query Match
 Best Local Similarity 96.7%; Score 29; DB 11; Length 321;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPMWR 5
 DB 44 FPMWR 48

RESULT 11
 O9APJ5 PRELIMINARY; PRT: 332 AA.
 AC O9APJ5;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DE ENDO-1,4-BETA-D-GLUCANASE PRECURSOR.
 GN CELBY.
 OS Erwiniia chrysanthemi.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Pectobacterium.
 OX NCBI_TaxID=556;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=PY35;
 RA Cho S.-J., Park S.-R., Yun H.-D.;
 RT "Cloning and sequencing of a celby gene of Pectobacterium chrysanthemi
 RT PY35."
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF282321; AAC49556.1;
 KW SIGNAL.
 FT SIGNAL 1 23 POTENTIAL.
 SO SEQUENCE 332 AA; 37627 MW; A1B507B0CB820EE6 CRC64;

Query Match
 Best Local Similarity 96.7%; Score 29; DB 2; Length 332;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPMWR 5
 DB 178 FPMWR 182

RESULT 12
 O9LJG9 PRELIMINARY; PRT: 333 AA.
 AC O9LJG9;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4).
 GN CELA.
 OS Erwiniia rhapontici.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Erwiniia.
 OX NCBI_TaxID=55212;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCPPB2989;
 RA Saarihahti H.T.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN 12
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCPPB2989;
 RA Riekl R.;
 RT "Members of the amylovora group of Erwiniia are cellulolytic and

RT possesses genes homologous to the type II secretion pathway."
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ276358; CAB89803.1;
 DR Interpro: IPR002037; Glyco_hydro_8.
 DR Pfam: PF01270; Glyco_hydro_8; 1.
 DR PRINTS: PR00735; GLYDCLASE8.
 KW SIGNAL: Hydrolase; Glycosidase.
 FT SIGNAL 1 23
 FT CHAIN 24 333
 SO SEQUENCE 333 AA; 37783 MW; D75CFE212302673A CRC64;

Query Match
 Best Local Similarity 96.7%; Score 29; DB 2; Length 333;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPMWR 5
 DB 178 FPMWR 182

RESULT 13
 O9FTM4 PRELIMINARY; PRT: 334 AA.
 AC O9FTM4;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DE P0005A05.25 PROTEIN.
 GN P0005A05.25.
 GN Oryza sativa (Rice).
 CS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 CC Eriacridaceae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
 RT clone: P0005A05."
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AP002863; BAB16921.1;
 DR Interpro: IPR000520; Exonuclease.
 DR Interpro: IPR000822; Znf-C2H2.
 DR Pfam: PF00929; Exonuclease 1.
 DR Pfam: PF00096; Zf-C2H2; 1.
 DR SMART: SM00479; EXO11; 1.
 DR SMART: SM00355; Znf-C2H2; 2.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
 DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 1.
 KW DNA-binding; Zinc-finger.
 SO SEQUENCE 334 AA; 37172 MW; A34531D625704A88 CRC64;

Query Match
 Best Local Similarity 96.7%; Score 29; DB 10; Length 334;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPMWR 5
 DB 301 FPMWR 305

RESULT 14
 O15429 PRELIMINARY; PRT: 349 AA.
 ID O15429;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE CD46.
 OS Homo sapiens (Human).

Search completed: February 27, 2002, 11:50:14
 Job time: 993 sec

OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;
 OC Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.
 OX NCBI_TaxID=9606;
 RN 11
 RP SEQUENCE FROM N.A.
 RC TISSUE=TESTIS;
 RA Hara T., Suzuki Y., Nagasawa S., Seya T.;
 RT "CDNA cloning and characterization of human sperm CD46.";
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: D84105; BAA12224.1; -
 DR HSPB: P10998; IYD;
 DR InterPro: IPR00436; Sushi_SCR_CCP.
 DR Pfam: PF00084; sushi; 4.
 DR SMART: SM00032; CCP; 4.
 SO SEQUENCE 349 AA; 39325 MW; 8FCEDA30D3C818E CRC64;

Query Match 96.7%; Score 29; DB 4; Length 349;
 Best Local Similarity 80.0%; Pred. NO. 4e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FPXWR 5
 11111
 DB 11 FPSWR 15

RESULT 15

O9SAAL PRELIMINARY: PRT: 417 AA.
 AC O9SAAL;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE F25C20.9.
 GN F25C20.9.
 CN F25C20.9.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
 OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae;
 OC eurosids II: Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
 RA Li J., Lee J.M., Kremetskaia I., Luos J., Ngan I., Liu A.,
 RA Gonzalez A., Altafi H., Araujo R., Chao Q., Conn L., Conway A.B.,
 RA Dunn P., Hansen N., Huizar L., Kim C., Palm C., Rowley D., Shinn P.,
 RA Walker M., Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
 RT "Arabidopsis thaliana chromosome 1 BAC F25C20 sequence.";
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN 12
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Theologis A.;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN 13
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Theologis;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC007296; AAC0247.1; -
 SO SEQUENCE 417 AA; 45985 MW; B86EBF06AC19A63D CRC64;

Query Match 96.7%; Score 29; DB 10; Length 417;
 Best Local Similarity 80.0%; Pred. NO. 4.7e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FPXWR 5
 11111
 DB 377 FPSWR 381

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OH protein - protein search, using sw model

Run on: February 27, 2002, 11:36:07 Search time 132.19 Seconds
(without alignments)
0.851 Million cell updates/sec

Title: US-09-446-109A-14

Perfect score: 30

Sequence: 1 FPXMR 5

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents AA:*

- 1: /cgn2-6/ptodata/2/1aa/5A-COMB.pep:*
- 2: /cgn2-6/ptodata/2/1aa/5B-COMB.pep:*
- 3: /cgn2-6/ptodata/2/1aa/6A-COMB.pep:*
- 4: /cgn2-6/ptodata/2/1aa/6B-COMB.pep:*
- 5: /cgn2-6/ptodata/2/1aa/PCTUS-COMB.pep:*
- 6: /cgn2-6/ptodata/2/1aa/backfilltest.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query length	DB ID	Description
1	29	96.7	38 1 US-08-176-500-118	Sequence 118, App
2	29	96.7	38 1 US-08-471-052A-118	Sequence 118, App
3	29	96.7	38 1 US-08-189-331-118	Sequence 118, App
4	29	96.7	38 2 US-08-471-939-118	Sequence 118, App
5	29	96.7	38 2 US-08-471-800-118	Sequence 118, App
6	29	96.7	38 2 US-08-471-068-118	Sequence 118, App
7	29	96.7	241 4 US-08-823-120-1	Sequence 1, Appl1
8	29	96.7	324 2 US-08-528-057-46	Sequence 46, Appl1
9	29	96.7	372 1 US-08-528-057-42	Sequence 42, Appl1
10	29	96.7	372 1 US-08-202-056-5	Sequence 5, Appl1
11	29	96.7	372 1 US-08-076-093A-6	Sequence 6, Appl1
12	29	96.7	372 1 US-08-701-265-6	Sequence 6, Appl1
13	29	96.7	372 2 US-08-284-586-6	Sequence 6, Appl1
14	29	96.7	372 2 US-08-805-478-6	Sequence 6, Appl1
15	29	96.7	372 2 US-08-802-627A-6	Sequence 6, Appl1
16	29	96.7	372 2 US-08-801-238-6	Sequence 6, Appl1
17	29	96.7	372 2 US-08-801-228-6	Sequence 6, Appl1
18	29	96.7	372 3 US-09-104-296-6	Sequence 6, Appl1
19	29	96.7	372 3 US-08-982-493-8	Sequence 6, Appl1
20	29	96.7	373 2 US-08-528-057-44	Sequence 8, Appl1
21	29	96.7	377 2 US-08-528-057-2	Sequence 44, Appl1
22	29	96.7	384 4 US-08-139-195-2	Sequence 2, Appl1
23	29	96.7	384 6 5514787-2	Patent No. 5514787
24	28	93.3	57 1 US-08-370-225-39	Sequence 29, Appl1
25	28	93.3	57 1 US-08-370-225-30	Sequence 30, Appl1
26	28	93.3	57 1 US-08-461-859-29	Sequence 29, Appl1
27	28	93.3	57 1 US-08-461-859-30	Sequence 30, Appl1

28	28	93.3	57 5 PCT-US93-10069-29	Sequence 29, Appl1
29	28	93.3	57 5 PCT-US93-10069-30	Sequence 30, Appl1
30	28	93.3	298 2 US-08-061-636-3	Sequence 3, Appl1
31	28	93.3	298 2 US-08-874-347-19	Sequence 19, Appl1
32	28	93.3	298 3 US-09-093-522-19	Sequence 19, Appl1
33	28	93.3	298 5 PCT-US94-05268-3	Sequence 13, Appl1
34	28	93.3	373 2 US-08-846-762-13	Sequence 6, Appl1
35	28	93.3	374 3 US-08-982-493-6	Sequence 1, Appl1
36	28	93.3	437 2 US-09-031-059-1	Sequence 3, Appl1
37	28	93.3	437 2 US-09-031-059-3	Sequence 2, Appl1
38	28	93.3	760 1 US-08-195-152-2	Sequence 2, Appl1
39	26	86.7	10 4 US-09-461-697-409	Sequence 409, App
40	26	86.7	12 4 US-09-461-697-407	Sequence 407, App
41	26	86.7	113 1 US-07-668-648-10	Sequence 10, Appl1
42	26	86.7	113 2 US-08-429-998-10	Sequence 10, Appl1
43	26	86.7	113 2 US-08-431-333-10	Sequence 10, Appl1
44	26	86.7	113 5 PCT-US91-02321-10	Sequence 10, Appl1
45	26	86.7	862 1 US-08-325-267A-4	Sequence 4, Appl1

ALIGNMENTS

RESULT 1
US-08-176-500-118
Sequence 118, Application US/08176500
Patent No. 5498538

GENERAL INFORMATION:
APPLICANT: Kay, B. K.
APPLICANT: Fowles, D. M.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 141
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/176,500
FILING DATE: 4/3/93
CLASSIFICATION: 435
PRIOR ART CITATION DATA:
APPLICATION NUMBER: US/08/013,416
FILING DATE: 4/3/93

ATTORNEY: INFORMATION:
NAME: Stock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-143
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-8090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-176-500-118

Query Match 96.7% Score 29; DB 1; Length 38;
Best Local Similarity 80.0% Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
11 11
Db 10 FPSMR 14

RESULT 2

US-08-471-052A-118
Sequence 118, Application US/08471052A
Patent No. 5625033
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 166
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471.052A
FILING DATE: 06-JUNE-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-179
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-471-052A-118

Query Match 96.7%; Score 29; DB 1; Length 38;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
11 11
Db 10 FPSMR 14

RESULT 3

US-08-189-331-118
Sequence 118, Application US/08189331
Patent No. 5747334
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 186
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/189,331
FILING DATE: Concurrently herewith
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-155
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-189-331-118

Query Match 96.7%; Score 29; DB 1; Length 38;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
11 11
Db 10 FPSMR 14

RESULT 4

US-08-471-939-118
Sequence 118, Application US/08471939
Patent No. 5844076
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 141
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471.939
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/013,416
FILING DATE: 01-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-143
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:

SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-471-939-118

Query Match 96.7% Score 29; DB 2; Length 38;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXWR 5
11 11
Db 10 FPSWR 14

RESULT 5
US-08-471-800-118
Sequence 118, Application US/08471800
Patent No. 5852167
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
APPLICANT: Fowlkes, D. M.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 141
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,800
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/013,416
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-143
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-471-800-118

Query Match 96.7% Score 29; DB 2; Length 38;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXWR 5
11 11
Db 10 FPSWR 14

RESULT 6
US-08-471-068-1
Sequence 118, Application US/08471068
Patent No. 5948635
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
APPLICANT: Fowlkes, D. M.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 186
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,068
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/189,331
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-155
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-471-068-118

Query Match 96.7% Score 29; DB 2; Length 38;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXWR 5
11 11
Db 10 FPSWR 14

RESULT 7
US-08-823-120-1
Sequence 1, Application US/08823120
Patent No. 6149919
GENERAL INFORMATION:
APPLICANT: Domenighini, Mario
APPLICANT: Rappuoli, Rino
TITLE OF INVENTION: Immunogenic Detoxified Mutants of
Cholera Toxin and of the Toxin B, their preparation and
their use for the preparation of vaccines
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: California
COUNTRY: USA

ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/823,120
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/256,003
FILING DATE: 11-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0315,001
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 241 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-823-120-1

Query Match 96.7% Score 29; DB 4; Length 241;
Best Local Similarity 80.0% Pred. No. 1,1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
11 11
Db 169 FPAWR 173

RESULT 8
US-08-528-057-46
Sequence 46 Application US/08528057
Patent No. 5846715
GENERAL INFORMATION:
APPLICANT: PURCELL, Damian F. J.
APPLICANT: RUSSELL, Sarah M.
APPLICANT: MCKENZIE, Ian F. C.
TITLE OF INVENTION: CD46 VARIANTS
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/528,057
FILING DATE: CONCURRENTLY HERewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/961,686
FILING DATE: 11-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU91/00199
FILING DATE: 10-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PK0133/90
FILING DATE: 11-MAY-1990

ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 17227/112 DACO
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 324 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-528-057-46

Query Match 96.7% Score 29; DB 2; Length 324;
Best Local Similarity 80.0% Pred. No. 1,4e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
11 11
Db 11 FPSWR 15

RESULT 9
US-08-528-057-42
Sequence 42 Application US/08528057
Patent No. 5846715
GENERAL INFORMATION:
APPLICANT: PURCELL, Damian F. J.
APPLICANT: RUSSELL, Sarah M.
APPLICANT: MCKENZIE, Ian F. C.
TITLE OF INVENTION: CD46 VARIANTS
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/528,057
FILING DATE: CONCURRENTLY HERewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/961,686
FILING DATE: 11-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU91/00199
FILING DATE: 10-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PK0133/90
FILING DATE: 11-MAY-1990
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 17227/112 DACO
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 370 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-528-057-42

Query Match 96.7%: Score 29; DB 2; Length 370;
Best Local Similarity 80.0%: Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPMR 5
DB 11 FPMR 15

RESULT 10
US-08-202-056-5
Sequence 5, Application US/08202056
Patent No. 5440021
GENERAL INFORMATION:
APPLICANT: Chuntharapai, Anan
APPLICANT: Hebert, Caroline
APPLICANT: Kim, Kyung Jin
APPLICANT: Lee, James
TITLE OF INVENTION: Antibodies to Human IL-8 Type B Receptor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Minipac (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202.056
FILING DATE: 25-FEB-1994
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/677211
FILING DATE: 29-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B.
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: 706P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: amino acid
TOPOLOGY: Linear
US-08-202-056-5

Query Match 96.7%: Score 29; DB 1; Length 372;
Best Local Similarity 80.0%: Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPMR 5
DB 352 FPMR 356

RESULT 11
US-08-076-093A-6
Sequence 6, Application US/08076093A
Patent No. 5543503
GENERAL INFORMATION:

APPLICANT: Chuntharapai, Anan
APPLICANT: Lee, James
APPLICANT: Hebert, Caroline
APPLICANT: Jin Kim, K.
TITLE OF INVENTION: Antibodies to Human PF4A Receptors
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Minipac (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/076.093A
FILING DATE: 11-Jun-1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/677211
FILING DATE: 29-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B.
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: 706P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-076-093A-6

Query Match 96.7%: Score 29; DB 1; Length 372;
Best Local Similarity 80.0%: Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPMR 5
DB 352 FPMR 356

RESULT 12
US-08-701-265-6
Sequence 6, Application US/08701265
Patent No. 5776457
GENERAL INFORMATION:
APPLICANT: Chuntharapai, Anan
APPLICANT: Lee, James
APPLICANT: Hebert, Caroline
APPLICANT: Jin Kim, K.
TITLE OF INVENTION: Antibodies to Human PF4A Receptors
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/701,265
FILING DATE: 22-AUG-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/076093
FILING DATE: 11-JUN-1993
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/677211
FILING DATE: 29-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: 706P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-701-265-6

Query Match 96.7%; Score 29; DB 1; Length 372;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FPXMR 5
11 11
Db 352 FPSWR 356

RESULT 13
US-08-284-586-6
Sequence 6, Application US/08284586
Patent No. 5840856
GENERAL INFORMATION:
APPLICANT: Chuntharapal, Anan
APPLICANT: Hebert, Caroline
APPLICANT: Jin Kim, K.
TITLE OF INVENTION: Antibodies to Human PFAA Receptors
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/284,586
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/076,093A
FILING DATE: 11-JUN-1993
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/677211
FILING DATE: 29-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: 706P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-284-586-6

Query Match 96.7%; Score 29; DB 2; Length 372;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FPXMR 5
11 11
Db 352 FPSWR 356

RESULT 14
US-08-805-478-6
Sequence 6, Application US/08805478
Patent No. 5874543
GENERAL INFORMATION:
APPLICANT: Chuntharapal, Anan
APPLICANT: Hebert, Caroline
APPLICANT: Jin Kim, K.
TITLE OF INVENTION: ANTIBODIES TO PFAA RECEPTOR
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/805,478
FILING DATE: 25-Feb-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/284586
FILING DATE: 10-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/076093
FILING DATE: 11-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B.
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: P0706P2P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:

LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-805-478-6

Search completed: February 27, 2002, 11:36:07
Job time: 146 sec

Query Match 96.7%; Score 29; DB 2; Length 372;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FPXMR 5
1111
Db 352 FPSMR 356

RESULT 15
US-08-802-627A-6
Sequence 6, Application US/08802627A
Patent No. 5892017
GENERAL INFORMATION:
APPLICANT: Lee, James
APPLICANT: Wood, William I.
TITLE OF INVENTION: NUCLEIC ACID ENCODING PFAA RECEPTOR
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 MB floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPacIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/802.627A
FILING DATE: 19-Feb-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/284586
FILING DATE: 10-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/076093
FILING DATE: 11-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B.
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: P0706P2P1D2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-802-627A-6

Query Match 96.7%; Score 29; DB 2; Length 372;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FPXMR 5
1111
Db 352 FPSMR 356

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:41:19 ; Search time 303.5 Seconds
(Without alignments)
1.220 Million cell updates/sec

Title: US-09-446-109a-15
Perfect score: 30
Sequence: 1 FPXMR 5

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /SIDSB/gcgdata/geneseq/AA1980.DAT:*
2: /SIDSB/gcgdata/geneseq/AA1981.DAT:*
3: /SIDSB/gcgdata/geneseq/AA1982.DAT:*
4: /SIDSB/gcgdata/geneseq/AA1983.DAT:*
5: /SIDSB/gcgdata/geneseq/AA1984.DAT:*
6: /SIDSB/gcgdata/geneseq/AA1985.DAT:*
7: /SIDSB/gcgdata/geneseq/AA1986.DAT:*
8: /SIDSB/gcgdata/geneseq/AA1987.DAT:*
9: /SIDSB/gcgdata/geneseq/AA1988.DAT:*
10: /SIDSB/gcgdata/geneseq/AA1989.DAT:*
11: /SIDSB/gcgdata/geneseq/AA1990.DAT:*
12: /SIDSB/gcgdata/geneseq/AA1991.DAT:*
13: /SIDSB/gcgdata/geneseq/AA1992.DAT:*
14: /SIDSB/gcgdata/geneseq/AA1993.DAT:*
15: /SIDSB/gcgdata/geneseq/AA1994.DAT:*
16: /SIDSB/gcgdata/geneseq/AA1995.DAT:*
17: /SIDSB/gcgdata/geneseq/AA1996.DAT:*
18: /SIDSB/gcgdata/geneseq/AA1997.DAT:*
19: /SIDSB/gcgdata/geneseq/AA1998.DAT:*
20: /SIDSB/gcgdata/geneseq/AA1999.DAT:*
21: /SIDSB/gcgdata/geneseq/AA2000.DAT:*
22: /SIDSB/gcgdata/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	96.7	10	20	AAV08336
2	29	96.7	11	20	AAV08334
3	29	96.7	38	15	AA858491
4	29	96.7	94	20	AAV06599
5	29	96.7	94	22	AA820130
6	29	96.7	135	20	AAV04918
7	29	96.7	177	21	AA858982
8	29	96.7	299	21	AA852488
9	29	96.7	314	17	AA83943
10	29	96.7	324	12	AA815233
11	29	96.7	359	8	AAV70396

12	29	96.7	372	13	AA827793	New platelet facto
13	29	96.7	372	16	AA892239	Chemokine superfam
14	29	96.7	372	16	AA868813	Human lymphocyte P
15	29	96.7	372	20	AAV06644	Human Burkitt's ly
16	29	96.7	372	21	AAV90627	Human G protein-co
17	29	96.7	372	12	AAV90661	Human G protein-co
18	29	96.7	373	11	AA815230	CD46 from clone pm
19	29	96.7	376	17	AA893939	CD46 wild-type. H
20	29	96.7	377	12	AA815229	CD46 from clone pm
21	29	96.7	377	17	AA893942	CD46 construct sub
22	29	96.7	377	17	AA893941	CD46 construct sub
23	29	96.7	377	18	AA827484	Human MCP. Homo s
24	29	96.7	384	12	AA810924	Human membrane cof
25	29	96.7	384	16	AA86316	Human CD46. Homo
26	29	96.7	418	21	AA813589	Streptomyces globi
27	29	96.7	421	21	AA858394	Lung cancer associ
28	29	96.7	421	22	AA875528	Human colon cancer
29	29	96.7	437	20	AAV04572	Mycobacterium spec
30	29	96.7	520	22	AA890739	C glutamicum prote
31	29	96.7	589	20	AAV34524	Porphyromonas ging
32	29	96.7	590	21	AAV94928	Human secreted pro
33	29	96.7	593	20	AAV34523	Porphyromonas ging
34	29	96.7	606	20	AAV34393	Porphyromonas ging
35	29	96.7	611	22	AA803762	CAB-2 chimeric pro
36	29	96.7	612	22	AA894498	Human protein sequ
37	29	96.7	614	18	AAW20991	H. pylori inner me
38	29	96.7	621	22	AA894049	Human protein sequ
39	29	96.7	634	22	AA840125	Human polypeptide
40	29	96.7	634	22	AA892742	Human protein sequ
41	29	96.7	634	22	AA893478	Human protein sequ
42	29	96.7	634	22	AA895120	Human protein sequ
43	29	96.7	1464	19	AAW79294	An antigen from de
44	29	93.3	42	28	AAW15108	Peptide #1542 enco
45	28	93.3	42	22	AAW27563	Peptide #1600 enco

ALIGNMENTS

RESULT 1

AAV08336 standard; Peptide: 10 AA.

AAV08336:

19-JUL-1999 (first entry)

O. hannah *gpcr* alpha-neurotoxin homologous peptide 5.

Snake; venom; King cobra; alpha-neurotoxin; toxin; analgesia; screening.

Ophiophagus hannah.

WO9924055-A1.

20-MAY-1999.

03-NOV-1998; 98WO-SC00087.

06-NOV-1997; 97SG-0003972.

(UYSI-) UNIV SINGAPORE NAT.

Gopaladrishtakone P, Gwee MCE, Klni RM, Pu XC, Wong PT;

WPI: 1999-327205/27.

Snake neurotoxin derived peptides

Claim 11: Page 43; 46pp; English.

This invention describes novel peptide derivatives of alpha-neurotoxin, derived from snake venom toxins of Ophiophagus hannah and which are

CC capable of inducing or facilitating analgesia. The peptides, optionally
 CC with a carrier and/or diluent, are used to induce analgesia in a subject.
 CC Antibodies derived from the products of the invention can be used to
 CC screen for subject peptides, e.g. for purification purposes.

XX
 SQ Sequence 10 AA;

Query Match

Best Local Similarity 96.7%; Score 29; DB 20; Length 10;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
 II II
 Db 3 fpwvr 7

RESULT 2

AAV08334
 ID AAV08334 standard; Peptide: 11 AA.

AC AAV08334;

XX 19-JUL-1999 (first entry)

DE O. hannah venom alpha-neurotoxin homologous peptide 3.

XX Snake; venom; King cobra; alpha-neurotoxin; toxin; analgesia; screening.

XX Ophiophagus hannah.

XX W09924055-A1.

XX 20-MAY-1999.

XX 03-NOV-1998; 98MO-SC00087.

XX 06-NOV-1997; 97SG-0003972.

XX (UYSI-) UNIV SINGAPORE NAT.

PI Copaladrishnakone P, Gwee MCE, Klini RM, Pu XC, Wong PT;

XX WPI: 1999-327205/27.

XX Snake neurotoxin derived peptides

PS Claim 9; Page 42; 46pp; English.

CC This invention describes novel peptide derivatives of alpha-neurotoxin,
 CC derived from snake venom toxins of Ophiophagus hannah and which are
 CC capable of inducing or facilitating analgesia. The peptides, optionally
 CC with a carrier and/or diluent, are used to induce analgesia in a subject.
 CC Antibodies derived from the products of the invention can be used to
 CC screen for subject peptides, e.g. for purification purposes.

XX
 SQ Sequence 11 AA;

Query Match 96.7%; Score 29; DB 20; Length 11;

Best Local Similarity 80.0%; Pred. No. 14;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
 II II
 Db 3 fpwvr 7

RESULT 3

AAV058491
 ID AAV058491 standard; Protein: 38 AA.

XX
 AC AAV058491;

XX 18-APR-1995 (first entry)

XX Binding domain of a polystyrene-binding TSAR (SB.9-5).

XX TSAR: totally synthetic affinity reagent; synthetic; binding domain;
 KW effector domain; concatenated heterofunctional protein; linker;
 KW direct; rapid; detection; screening; treatment; monoclonal antibody;
 KW Sm antigen; mouse model; autoimmune disease; SLE; Mab;
 KW systemic lupus erythematosus; polystyrene; SB.

XX Synthetic.

XX W09418318-A.

XX 18-AUG-1994.

XX 01-FEB-1994; 94MO-US00977.

XX 01-FEB-1993; 93US-0013416.

XX 30-DEC-1993; 93US-0176500.

XX 31-JAN-1994; 94US-0189331.

XX (UYNC-) UNIV NORTH CAROLINA.

XX Fowlkes DM, Kay BK;

XX WPI: 1994-279739/34.

XX Identifying proteins or peptide(s) which bind a ligand - by
 PT screening a recombinant vector library expressing fusion proteins
 PT comprising a binding domain and an effector domain

XX Example 7.8; Page 120; 255pp; English.

CC AAV58491-507 are amino acid sequences of the binding domain of TSAR
 CC (Totally Synthetic Affinity Reagents) peptides. These particular
 CC peptides are encoded by the polystyrene (SB) binding phage. There
 CC is no apparent linear motif, but the peptides are rich in Trp, Tyr and
 CC Gly, poor in Arg, Val and Lys, and completely lack cysteine residues.
 CC The non-variable amino acids at the NH2 and COOH terminals are
 CC not shown. TSAR peptides are generated using generic oligonucleotides
 CC (see AAV070470-73 for examples). TSARs are concatenated heterofunctional
 CC proteins or peptides, comprising at least two functional regions - a
 CC binding domain with affinity for a ligand and a second effector peptide
 CC portion that is chemically or biologically active. They may further
 CC comprise a linker peptide between the 2 domains. The TSARs or compns.
 CC comprising a TSAR binding domain can be used in vivo to deliver a
 CC chemically or biologically active moiety, eg. metal ion, radioisotope,
 CC peptide, toxin or enzyme, to the specific target or on the cell. They
 CC can also replace the function of macromolecules eg. monoclonal or
 CC polyclonal antibodies and therefore circumvent the need for complex
 CC methods of hybridoma formation or in vivo antibody production. The TSARs
 CC are easily characterised and have designed actively allowing direct and
 CC rapid detection in a screening process.

XX
 SQ Sequence 38 AA;

Query Match 96.7%; Score 29; DB 15; Length 38;

Best Local Similarity 80.0%; Pred. No. 46;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
 II II
 Db 10 fpwvr 14

RESULT 4

AAV06599
 ID AAV06599 standard; Protein: 94 AA.

XX
 AC AAV06599;

[illegible]

PN MO200105831-AL.
XX
PD 25-JAN-2001.
XX
PF 14-JUL-2000; 2000MO-USI9342.
XX
PR 16-JUL-1999; 99US-0354147.
XX
PA (UYVA) UNIV YALE.
XX
PI DiB-Hajj S, Waxman SG;
XX WPI: 2001-103147/11.
DR
XX Nucleic acid molecules encoding human tetrodotoxin resistant sodium
PT channels, useful for preventing, diagnosing and treating pain,
FT paraesthesia and/or hyperexcitability phenomena -
XX
XS Example 5; Page 34; 162pp; English.
XX
CC The present sequence is that of the C-terminal portion of a
CC truncated rat Nan variant. The variant lacks the 387 C-terminal
CC amino acids of full-length rat Nan (see AAB20122). The gene is
CC replaced by the present 94-amino acid polypeptide. The N-terminal
CC 1378 amino acids of the truncated variant are identical to those
CC of full-length Nan. The new sequence arises from the use of a
CC cryptic donor splice site in exon 23 and a novel exon 23 which is
CC located in intron 23. Rat Nan is a novel tetrodotoxin resistant
CC sodium channel Nan belonging to the a-subunit voltage-gated sodium
CC channel protein family. These sodium channels underlie the
CC generation and propagation of impulses in excitable cells such as
CC neurons and muscle fibres. Preferential expression of Nan on
CC sensory neurons innervating the body (dorsal root ganglia) and
CC the face (trigeminal ganglia), but not on other neurons, makes
CC it a very useful target for diagnostic and/or therapeutic uses in
CC relation to acute and/or chronic pain pathologies. A claim-
CC method of treating pain, paraesthesia and/or hyperexcitability
CC phenomena in a human or animal subject involves administering an
CC agent that alters sodium current flow through Nan channels, or
CC which modulates transcription or translation of Nan mRNA, in
CC dorsal root ganglia or trigeminal neurons. Nan polypeptides can
CC be obtained by recombinant expression, and used to treat disorders
CC associated with decreased sodium channel expression, to screen for
CC compounds that modulate sodium channel expression or activity,
CC and to raise antibodies useful as diagnostic agents.
SQ Sequence 94 AA:

Query Match 96.7%; Score 29; DB 22; Length 94;
Best Local Similarity 80.0%; Pred.No. 1.le+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

VY 1 PFXMR 5
II II
Db 52 fpawr 56

RESULT 6
AAV04918
ID AAV04918 standard: Protein; 135 AA.
XX
AC AAV04918;
XX
UT 06-JUL-1999 (first entry)
XX
NE Mycobacterium species protein sequence 35B.
XX
KW Secreted protein; Mycobacterium; primer: PCR: amplification; probe:
XX hybridisation; detection: vaccine; immunisation; infection.
OS Mycobacterium sp.
XX

PN W09909186-A2.
 XX
 PD 25-FEB-1999.
 XX
 PE 14-AUG-1998; 98WO-FR01813.
 XX
 PR 11-SEP-1997; 97FR-0011325.
 PR 14-AUG-1997; 97FR-0010404.
 XX
 PA (INST) INST PASTEUR.
 XX
 PI Gicquel B, Lim EM, Pelicic V, Portnoi D, Coquet de la Salmoniere Y;
 PI Guigueno A.
 XX
 DR WPI: 1999-181045/15.
 DR N-PSDB: AAX341171.
 XX
 PT Mycobacterial DNA vectors containing reporter constructs - for
 PT Identifying coding or promoter sequences involved in
 PT Infection-associated protein expression
 XX
 PS Claim 32: Fig 35B; 309pp; French.
 XX
 CC Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted
 CC proteins from various Mycobacterium species microorganisms. The
 CC encoding nucleotide sequences can be used as primers and probes for
 CC methods for detecting and identifying mycobacteria, especially belonging
 CC to the M. tuberculosis complex. The encoded proteins can be used in
 CC vaccines for immunisation against a bacterial or viral infection.
 XX
 SQ Sequence 135 AA;

Query Match 96.7%; Score 29; DB 20; Length 135;
 Best Local Similarity 80.0%; Pred. No. 1.5e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
 11 11
 Db 66 fpwvr 70

RESULT 7
 AAB58892
 ID AAB58892 standard; Protein: 177 AA.

AC AAB58892;

DT 27-MAR-2001 (first entry)

DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 600.

XX Human: breast cancer; cytostatic; immunosuppressive;
 KW neotropic; neurprotection; antiviral; antiallergic; hepatotropic;
 KW antidiabetic; antinflammatory; antitumor; vulnereary; anticonvulsant;
 KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
 KW Addison's disease; allergy; autoimmune haemolytic anaemia;
 KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
 KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
 KW cardiovascular disorder; wound healing; neurological disease.

XX Homo sapiens.

XX W0200055173-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05881.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX

PI Rosen CA, Ruben SM;
 XX
 DR WPI: 2000-611515/58.
 DR N-PSDB: AAF21795.
 XX
 PT New human breast and ovarian cancer associated gene sequences and the
 PT polypeptides encoded by these genes, useful in the prevention,
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular
 PT disorders and neurological diseases -
 XX
 PS Claim 11: Page 1037-1038; 1299pp; English.

XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
 CC associated with breast and ovarian cancer. Included in the invention are
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
 CC isolation and characterisation of the DNA and protein sequences of the
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist
 CC or antagonist sequences exhibit cytostatic; immunosuppressive;
 CC neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 CC antidiabetic; antinflammatory; antitumor; vulnereary; anticonvulsant;
 CC polynucleotide and protein sequences are used in the diagnosis of cancer,
 CC particularly breast and ovarian cancer. The nucleic acid sequences,
 CC proteins, agonists and antagonists may also be used in the diagnosis,
 CC prevention and treatment of immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; cardiovascular disorders such as
 CC myocardial ischaemias; wound healing; neurological diseases such as
 CC cerebral anoxia and epilepsy; and infectious diseases.
 XX
 SQ Sequence 177 AA;

Query Match 96.7%; Score 29; DB 21; Length 177;
 Best Local Similarity 80.0%; Pred. No. 2e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
 11 11
 Db 22 fpwvr 26

RESULT 8

AAB52488
 ID AAB52488 standard; Protein: 299 AA.

AC AAB52488;

DT 23-FEB-2001 (first entry)

DE Helicobacter pylori bait polypeptide #6.

XX Helicobacter pylori; two-hybrid system; protein-protein interaction;
 KW bait polypeptide; gastric ulcer; antibacterial.

XX Helicobacter pylori.

XX W0200066722-A1.

XX 09-NOV-2000.

XX 14-APR-2000; 2000WO-IB00603.

XX 30-APR-1999; 99EP-0401066.

XX (HYBR-) HYBRIGENICS SA.

XX Legrain P, Selig L, Rain J;

XX WPI: 2000-687535/67.

XX N-PSDB: AAC97234.

XX

```

XX A two-hybrid system for identifying compounds useful in the treatment
PT of e.g. gastric ulcers comprises producing a collection of recombinant
PT cell clones -
XX
PS Example 5: Page 88-89; 267pp; English.
XX
CC The present sequence is a bait polypeptide used in a Helicobacter
CC pylori two-hybrid screen to identify protein-protein interactions.
CC The method is used to identify a recombinant cell clone expressing a
CC prey polypeptide which is capable of interacting with the bait
CC polypeptide. The two hybrid system is useful for screening compounds
CC for antibacterial activity. It may be used in the treatment of gastric
CC ulcers. The polynucleotides are useful as amplification primers or
CC specific detection probes. The polypeptides, vectors or host cells can
CC be used as immunogens to produce mono- or polyclonal antibodies. The
CC polynucleotides, polypeptides, antibodies, vectors, host cells or
CC modulating agents can be used to produce a pharmaceutical composition.
XX
SQ Sequence 299 AA:
XX
Query Match 96.7%; Score 29; DB 21; Length 299;
Best Local Similarity 80.0%; Pred. No. 3.3e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 FPXMR 5
Db 89 fpsvr 93
XX
RESULT 9
AAR93943
ID AAR93943 standard; Protein: 314 AA.
XX
AAR93943:
XX
21-MAY-1996 (first entry)
XX
CD46 construct delSCR2/subSCR3.
XX
CD46; recombinant protein; short consensus repeat; SCR:
KW regulator of complement activation; transgenic animal; pig:
KM organ transplantation.
XX
Synthetic.
XX
WO9606937-A1.
XX
07-MAR-1996.
XX
30-AUG-1995; 95WO-AU00553.
XX
30-AUG-1994; 94AU-0007724.
XX
(AUST-) AUSTIN RES INST.
XX
Christiansen D, Loveland B, McKenzie IFC, Milland J;
XX
WPI: 1996-160368/16.
XX
N-PSDB: AAT17599.
XX
Increasing prodn. of recombinant proteins, esp. CD46 - by reducing
PT the amt. of A and/or T in an A and/or T rich region of encoding gene
PT exon
XX
Claim 12: Page 40-41; 60pp; English.
XX
CD46 delSCR2/subSCR3 (AAR93943) is the product of a cDNA construct
CC (AAT17599) obtd. by splice overlap extension PCR of wild-type CD46
CC cDNA (AAT17595). The A+T content of A+T-rich exon 5 of the gene,
CC encoding the short consensus repeat (SCR) 3 of CD46, was lowered
CC and SCR2 was deleted. This resulted in improved prodn. in

```

```

CC eukaryotic host cells, e.g. CHO-K1 and COS-7. CD46 delSCR2/subSCR3
CC is used to prevent complement- or inflammation-mediated tissue
CC damage, to improve immunity to tumours or viruses, to control
CC fertilisation and to prevent spontaneous abortion. Expression in
CC transgenic animals, esp. pigs, provides organs suitable for
CC transplantation.
XX
SQ Sequence 314 AA:
XX
Query Match 96.7%; Score 29; DB 17; Length 314;
Best Local Similarity 80.0%; Pred. No. 3.4e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 FPXMR 5
Db 11 fpsvr 15
XX
RESULT 10
AAR15233
ID AAR15233 standard; Protein: 324 AA.
XX
AAR15233:
XX
17-MAR-1992 (first entry)
XX
CD46 from clone pm5.8.
XX
SCR; short consensus repeats; transmembrane; cytoplasmic;
KW membrane cofactor protein; MCP.
XX
Homo sapiens.
XX
Key Location/Qualifiers
XX Peptide 1..34
XX Protein /label= sig-peptide
XX /label= 34..324
XX /label= mat-protein
XX Modified-site 83
XX /label= N-glycosylation_site
XX Modified-site 114
XX /label= N-glycosylation_site
XX Domain 289..304
XX /label= hydrophobic_transmembrane_domain
XX
WO9118097-A.
XX
28-NOV-1991.
XX
10-MAY-1991; 91WO-AU00199.
XX
11-MAY-1990; 90AU-0000133.
XX
(WYNE-) UNIV MELBOURNE.
XX
Purcell DFJ, Russell SM, McKenzie JFC;
XX
WPI: 1991-369251/50.
XX
N-PSDB: AAQ14919.
XX
New CD46 membrane co-factor protein variants - useful as probes
PT to identify CD46 isoforms and for diagnosing spontaneous
PT abortion, inhibiting immuno:response and treating leukaemia
XX
Disclosure: Fig 1A and 3B; 77pp; English.
XX
The sequence of the pm5.8 clone is identical to the pm5.1 clone in
CC the portion encoding the NH2 leader and four SCR regions. The
CC sequence after nucleotide 890 was different, and results from
CC reading through of the cDNA into an intron sequence after the
CC fourth SCR. The protein encoded by this new sequence encodes
CC a 16 amino acid hydrophobic region, probably serving as a membrane

```

CC spanning sequence. The putative cytoplasmic tail would also
 CC contain a high proportion of charged amino acids.
 CC See also AAQ14915-25, AAQ15211-12 and AAR15457-59.
 XX

SQ Sequence 324 AA;

Query Match 96.7%; Score 29; DB 12; Length 324;
 Best Local Similarity 80.0%; Pred. No. 3.5e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
 11 fpswr 15

RESULT 11

AAQ70396 standard; Protein; 359 AA.

XX
 XX
 AC AAR70396;
 XX
 DT 30-APR-1991 (first entry)
 XX
 DE Cellulase.
 XX
 KM Cellulase; plasmid; transformation.
 XX
 OS Cellulomonas uda CB4 (FER.
 XX
 PN JP62175178-A.
 XX
 PD 31-JUL-1987.
 XX
 PF 30-JAN-1986; 86JP-0016800.
 XX
 PR 30-JAN-1986; 86JP-0016800.
 XX
 PA (ENER-) SHIN ENERGY SOGO KAIHATSU KIKO.
 XX
 PA (NENR-) NENRYOYO ALCOHOL KAIHATSU GIJUTSU KENKYU KUMIAT.
 XX
 DR WPI: 1987-253827/36.
 DR N-PSDB: AAN70617.
 XX
 XX

PT Bio-engineering cellulase prodn. - by preparing DNA chain,
 PT preparing plasmid conty. DNA chain in the gene expressible state,
 PT and transforming bacterial by plasmid

PS Disclosure; Fig 1(A-B); 12pp; Japanese.

CC Transformation E.coli JM103 (FERM P-8612) with the gene encoding
 CC this protein allows expression of the protein and easy and cheap
 CC prodn. of cellulase.
 CC
 XX

SQ Sequence 359 AA;

Query Match 96.7%; Score 29; DB 8; Length 359;
 Best Local Similarity 80.0%; Pred. No. 3.9e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
 11 fpswr 182

RESULT 12

AAQ27793 standard; Protein; 372 AA.

XX
 XX
 AC AAR27793;
 XX
 DT 12-MAR-1993 (first entry)

XX
 DE New platelet factor 4 receptor superfamily member PF4AR11.
 XX
 XX
 KM IL-8R: G-protein coupled receptor family; rhodopsin superfamily;
 KM pro-inflammatory cytokine; 8rr.9.
 XX
 XX

OS Homo sapiens.

PN W09217497-A.

PD 15-OCT-1992.

PF 23-MAR-1992; 92MO-US02317.

PR 29-MAR-1991; 91US-0677211.

PR 19-DEC-1991; 91US-0810782.

PA (GETH) GENENTECH INC.

PI Holmes WE, Lee J, Wood WI;

PI WPI: 1992-366191/44.

PI N-PSDB: AAQ37107.

PT Isolated human platelet factor 4 super-family receptor

PT polypeptide and corresp. antibodies and DNA - useful as

PT diagnostic and screening agents, and for treating inflammation or

PT PF4AR-mediated disorders

XX
 XX
 TS Claim 7; Fig 5; 78pp; English.

CC The IL-8 receptor cDNA sequence was isolated (see AAQ29505) and a
 CC 874bp sub-fragment of the coding sequence was used as a probe to
 CC screen human cell line HL60 and human peripheral blood lymphocyte
 CC cDNA libraries. Two new gene sequences were found that are clearly
 CC related to the IL-8 receptor. One of these was contained in clone
 CC 8rr.9 and is predicted to encode an amino acid sequence which is
 CC 36% and 38% identical with the high and low affinity IL-8 receptor
 CC sequences, respectively. See also AAQ37107.
 XX
 XX

SQ Sequence 372 AA;

Query Match 96.7%; Score 29; DB 13; Length 372;
 Best Local Similarity 80.0%; Pred. No. 4e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
 11 fpswr 356

RESULT 13

AAQ92239 standard; Protein; 372 AA.

XX
 XX
 AC AAR92239;
 XX
 DT 26-MAR-1996 (first entry)

DE Chemokine superfamily receptor.

KM Interleukin; IL-8; inflammation; psoriasis; dermatitis;

KM rheumatoid arthritis; inflammatory bowel disease;

KW chronic lung inflammation; treatment; antibody;

KW affinity purification; detection.

OS Homo sapiens.

PN US5440021-A.

XX
 XX
 PD 08-AUG-1995.
 XX

PF 29-MAR-1991: 91US-0677211.
 XX
 XX 25-FEB-1994: 94US-0202056.
 PR 29-MAR-1991: 91US-0677211.
 XX
 XX (CHUN/) CHUNTHARAPAI A.
 PA (HEBE/) HEBERT C.
 PA (KIM/) KIM K J.
 PA (LEE/) LEE J.
 XX
 PI Chuntharapai A, Hebert C, Kim KJ, Lee J;
 XX
 DR WPI: 1995-283151/37.
 DR N-PSDB: AAQ99009.
 XX
 PT New antibodies against Interleukin 8 type B receptor - used to treat
 PT or prevent inflammation, also for detecting receptor expression and
 PT purification.
 PS
 PS Example 2: Columns 49-52; 62pp: English.
 XX
 CC Antibodies directed against the interleukin-8 receptor B can be used
 CC to treat or prevent inflammation e.g. psoriasis, dermatitis,
 CC rheumatoid arthritis and particularly inflammatory bowel disease and
 CC chronic lung inflammation. When immobilised, these antibodies may
 CC be used to detect interleukin-8 receptor B expression in cells and
 CC tissues and for affinity purification of interleukin-8 receptor B
 CC from cells. This sequence is an additional chemokine superfamily
 CC receptor which was identified by probing lambda libraries of genomic
 CC DNA from a human monocyte-like cell line (U-60) and human peripheral
 CC blood lymphocytes using a large fragment of the interleukin-8 type
 CC A receptor DNA (See AAQ99006).
 CC
 SQ Sequence 372 AA:
 QY 1 FPXWR 5
 1111
 DB 352 fpswr 356
 OY
 DB 352 fpswr 356
 RESULT 14
 ID AAR68813 standard: Protein: 372 AA.
 XX
 AC AAR68813:
 XX
 DT 18-JUL-1995 (first entry)
 XX
 DE Human lymphocyte PFAAR.
 XX
 KW Interleukin-8 receptor; IL-8 receptor; PFAAR;
 KW Platelet factor superfamily receptor; Lymphocyte; chemotactic;
 KM Inflammation; inflammatory disease; arthritis; emphysema; cystic;
 KM fibrosis; colitis; bronchitis; meningitis; therapeutic.
 XX
 OS Homo sapiens.
 XX
 PN MO9428931-A.
 XX
 PD 22-DEC-1994.
 XX
 XX 07-JUN-1994: 94WO-US06380.
 XX
 PK 11-JUN-1993: 93US-0076093.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Chuntharapai A, Hebert C, Kim KJ, Lee J;

XX
 DR WPI: 1995-036114/05.
 PR N-PSDB: AAQ80522.
 XX
 PT Treatment of inflammatory disorders - by administering an
 PT antibody capable of binding a platelet factor 4 superfamily
 PT receptor polypeptide
 XX
 PS Disclosure: Page 56-58; 83pp: English.
 XX
 CC 2 PFAAR members were identified by probing lambda libraries from
 CC human monocyte-like cell line HL-60 and human peripheral blood
 CC lymphocytes using a large fragment of IL-8 receptor DNA (full
 CC sequence given in AAQ80520). The nucleotide sequences of the 2
 CC PFAARs are given in AAQ80521 and AAQ80522, and their respective
 CC amino acid sequences in AAR68812 and AAR68813.
 CC
 SQ Sequence 372 AA:
 QY 1 FPXWR 5
 1111
 DB 352 fpswr 356
 OY
 DB 352 fpswr 356
 RESULT 15
 ID AAY06644 standard: Protein: 372 AA.
 XX
 AC AAY06644:
 XX
 DT 26-OCT-1999 (first entry)
 XX
 DE Human Burkitt's lymphoma receptor 1 (BLR1).
 XX
 KW Burkitt's lymphoma receptor 1; BLR1; human;
 KW B lymphocyte chemottractant; BLC; chemokine; ligand;
 KW drug screening; leukemia; autoimmune disease; therapy.
 XX
 OS Homo sapiens.
 XX
 PN MO9928468-A1.
 XX
 PD 10-JUN-1999.
 XX
 PF 02-DEC-1998: 98WO-US25561.
 XX
 PR 02-DEC-1997: 97US-0982493.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Cyster JC, Gunn MD, Williams LT;
 XX
 DR WPI: 1999-493764/41.
 DR N-PSDB: AAX87710.
 XX
 PT Modulating interaction of a Burkitt's Lymphoma Receptor 1
 PT polypeptide and ligand, useful in drug screens
 PS
 PS Claim 1: Page 38; 42pp: English.
 XX
 CC This sequence represents human Burkitt's lymphoma receptor 1
 CC (BLR1). The invention relates to methods for modulating the
 CC interaction of BLR1 with its ligand, B lymphocyte chemottractant
 CC (BLC, see AAY06642). The methods comprise combining BLR1 and BLC
 CC polypeptides with a candidate modulator agent under conditions
 CC whereby, but for the presence of the agent, the polypeptides engage
 CC in a first interaction, and determining a second interaction of the
 CC polypeptides in the presence of the agent, wherein a difference

CC between the first and second interactions indicates that the agent
CC modulates the interaction of the polypeptides. The modulator is
CC preferably an antagonist, especially dominant negative, form of
CC BLC. BLC and BLC agonists and antagonists may be useful for
CC treating viral (e.g. HIV) infections, lymphoma, B lineage
CC leukemia, and autoimmune diseases such as rheumatoid arthritis,
CC thyroiditis and diabetes.
XX
SQ Sequence 372 AA;

Query Match: 96.7%; Score 29; DB 20; Length 372;
Best Local Similarity 80.0%; Pred. No. 4e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXWR 5
1111
Db 352 fpswr 356

Search completed: February 27, 2002, 11:41:19
Job time: 458 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:45:23 : Search time 145.23 seconds
(without alignments)
2.623 Million cell updates/sec

Title: US-09-446-109a-15

Perfect score: 30

Sequence: 1 PPXWR 5

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database:

1: PIR68:*
2: PIR1:*
3: PIR2:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
1	29	96.7	73	1	N20H1	long neurotoxin 1
2	29	96.7	113	2	T07855	translational elonga
3	29	96.7	157	2	T24380	hypothetical prote
4	29	96.7	245	2	D82680	conserved hypothet
5	29	96.7	256	2	D85510	probable EC 3.5. a
6	29	96.7	259	2	A29831	heat-labile entero
7	29	96.7	273	2	T05454	hypothetical prote
8	29	96.7	292	2	G82960	hypothetical prote
9	29	96.7	327	2	S56162	MDCR15 protein - h
10	29	96.7	332	2	JT0585	minor endoglucanas
11	29	96.7	349	2	G02913	sperm CD46 - human
12	29	96.7	359	2	I40696	endoglucanase - Ce
13	29	96.7	369	2	I57998	membrane cofactor
14	29	96.7	372	2	S26667	G protein-coupled
15	29	96.7	377	2	I54479	membrane cofactor
16	29	96.7	384	2	S01896	membrane cofactor
17	29	96.7	417	2	D86251	protein F25C20.9 l
18	29	96.7	479	2	T35441	aldehyde dehydroge
19	29	96.7	497	2	D83528	probable aldehyde
20	29	96.7	555	2	F69312	heterodisulfide re
21	29	96.7	625	2	H70330	hypothetical prote
22	29	96.7	642	2	D71909	ferrous iron trans
23	29	96.7	642	2	G64605	iron(II) transport
24	29	96.7	650	2	T22002	hypothetical prote
25	29	96.7	735	2	A83006	hypothetical prote
26	29	96.7	1451	2	I40325	dermoecrotic toxin
27	28	93.3	68	2	S75058	transposase ssr289
28	28	93.3	84	2	S76091	hypothetical prote
29	28	93.3	84	2	S76443	hypothetical prote

30	28	93.3	93	2	S75008	transposase ss1192
31	28	93.3	103	2	T47718	hypothetical prote
32	28	93.3	114	2	S77061	transposase sl1066
33	28	93.3	143	1	HHK48	heat shock protein
34	28	93.3	143	1	HHK41	heat shock protein
35	28	93.3	147	2	A71560	hypothetical prote
36	28	93.3	149	2	T35846	probable integral
37	28	93.3	157	2	F81710	conserved hypothet
38	28	93.3	169	2	G69300	conserved hypothet
39	28	93.3	254	2	S76459	transposase sl1186
40	28	93.3	254	2	E82791	conserved hypothet
41	28	93.3	258	2	S45682	acetone-cyanhydrin
42	28	93.3	259	2	S76643	transposase slr051
43	28	93.3	261	2	S77171	transposase sl1171
44	28	93.3	261	2	S77351	transposase sl1171
45	28	93.3	261	2	S75081	transposase slr026

ALIGNMENTS

RESULT 1
N20H1
long neurotoxin 1 - king cobra
N:Alternate names: neurotoxin A
C:Species: Ophiophagus hannah (king cobra)
C:Date: 22-Jun-1981 #sequence_revision 22-Jun-1981 #text_change 23-Aug-1996
C:Accession: A01658
R:Joubert, F.J.
Biochim. Biophys. Acta 317, 85-98, 1973
A:Title: Snake venom toxins. The amino acid sequences of two toxins from Ophiophagus
A:Reference number: A90588; MUID:7321298
A:Accession: A01658
A:Molecule type: protein
A:Residues: 1-73 <300>
C:Superfamily: snake toxin
C:Keywords: neurotoxin; venom
F:3-21,14-42,27-31,46-57,58-63/Disulfide bonds: #status predicted

Query Match 96.7% Score 29: DB 1: Length 73:
Best Local Similarity 80.0% Pred. No. 33:
Matches 4: Conservative 0: Mismatches 1: Indels 0: Caps 0:

QY 1 PPXWR 5
DB 66 FPMWR 70

RESULT 2
T07855
translational elongation factor eEF-1 alpha chain - rape (fragment)
N:Alternate names: translation elongation factor EF-1A
C:Species: Brassica napus (rape)
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 31-Mar-2000
C:Accession: T07855
R:Saez-Vasquez, J.
Submitted to the EMBL Data Library, February 1995
A:Reference number: Z16177
A:Accession: T07855
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-113 <SAB>
A:Cross-references: EMBL:021744; NID:g914912; PIDN:AAA86366.1; PID:g914913
A:Experimental source: cv. Samourai; 5 days old etiolated seedlings
C:Superfamily: translation elongation factor Tu; translation elongation factor Tu hom
C:Keywords: GTP binding; protein biosynthesis

Query Match 96.7% Score 29: DB 2: Length 113:
Best Local Similarity 80.0% Pred. No. 50:
Matches 4: Conservative 0: Mismatches 1: Indels 0: Caps 0:

Qy 1 FPXMR 5
Db 103 FPMWR 107

RESULT 3

T24380
hypothetical protein T03D8.2 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24380
R:Morimoto, B.
submitted to the EMBL Data Library, March 1997
A:Reference number: Z19884
A:Accession: T24380
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1157 <MIL>
A:Cross-references: EMBL:Z92838; PIDN:CAB07406.1; GSPDB:GN00023; CESP:T03D8.2
A:Experimental source: clone T03D8
C:Genetics:
A:Gene: CESP:T03D8.2
A:Map position: 5
A:Introns: 49/3; 83/1; 125/3

Query Match 96.7%; Score 29; DB 2; Length 157;

Best Local Similarity 80.0%; Pred. No. 68;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FPXMR 5
Db 40 FPMWR 44

RESULT 4

D82680
conserved hypothetical protein XF1438 [imported] - *Xylella fastidiosa* (strain 9a5c)
C:Species: *Xylella fastidiosa*
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: D82680
R:Anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
A:Reference number: A82515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: D82680
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-245 <SIM>
A:Cross-references: GB:AE003974; GB:AE003849; NID:9106454; PIDN:AAF84247.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Simpton, A.J.G.; Relnach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briñon, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir, I
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
Submitted to Genbank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krueger, J.E.; Kuramae, E.E.; Laig
Chado, M.A.; Madlira, A.M.B.N.; Medeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, H
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1438

Query Match 96.7%; Score 29; DB 2; Length 245;
Best Local Similarity 80.0%; Pred. No. 1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FPXMR 5
Db 197 FPMWR 201

RESULT 5

D85510
probable EC 3.5. amidase-type enzyme yafV [imported] - *Escherichia coli* (strain O157:
C:Species: *Escherichia coli*
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: D85510
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: D85510
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-256 <STO>
A:Cross-references: GB:AE005174; NID:912512987; PIDN:AG54544.1; GSPDB:GN00145; UMCP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: yafV

Query Match 96.7%; Score 29; DB 2; Length 256;

Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FPXMR 5
Db 244 FPMWR 248

RESULT 6

A29831
heat-labile enterotoxin Iia chain A precursor - *Escherichia coli*
R:Alternate names: LT-Iia
C:Species: *Escherichia coli*
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C:Accession: A29831
R:Pickett, C.L.; Weinstein, D.L.; Holmes, R.K.
J. Bacteriol. 169, 5180-5187, 1987
A:Title: Genetics of type Iia heat-labile enterotoxin of *Escherichia coli*: operon *fus*
A:Reference number: A91849; MUID:88032841
A:Accession: A29831
A:Molecule type: DNA
A:Residues: 1-259 <PIC>
A:Cross-references: GB:M17894; NID:9146671; PIDN:AA24093.1; PID:9146672
A:Note: the authors translated the codon TAT for residue 225 as Thr
F:118/Domain: signal sequence #status predicted <SIG>
F:19-259/Product: heat-labile enterotoxin Iia chain A #status predicted <LTA>

Query Match 96.7%; Score 29; DB 2; Length 259;

Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FPXMR 5
Db 187 FPMWR 191

RESULT 7

T05454
hypothetical protein F7K2.180 - *Arabidopsis thaliana*
C:Species: *Arabidopsis thaliana* (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Sep-1999
C:Accession: T05454
R:Bevan, M.; Wedler, H.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schue

submitted to the Protein Sequence Database, November 1998
A:Reference number: 215416

A:Accession: T05454

A:Molecule type: DNA

A:Residues: 1-273 <REV>

A:Cross-references: EMBL:AL033545

A:Experimental source: cultivar Columbia; BAC clone F7K2

C:Genetics:

A:Map position: 4

A:Note: F7K2.180

C:Superfamily: Arabidopsis hypothetical protein F7K2.180

Query Match 96.7%: Score 29; DB 2; Length 273;
Best Local Similarity 80.0%: Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
1111
Db 79 FPSMR 83

RESULT 8

G82960 hypothetical protein PA5488 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: G82960

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardly, K.; Lam, J.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen

A:Reference number: AB2950; MUID:20437337

A:Accession: G82960

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-292 <STO>

A:Cross-references: GB:AE004961; GB:AE004091; NID:g9951814; PIDN:AAG08873.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA5488

Query Match 96.7%: Score 29; DB 2; Length 292;
Best Local Similarity 80.0%: Pred. No. 1.2e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
1111
Db 231 FPSMR 235

RESULT 9

S56162 MOCR15 protein - human

C:Species: Homo sapiens (man)

C>Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999

C:Accession: S56162

R:Barclite, L.; Loetscher, M.; Tobler, A.; Baggiolini, M.; Moser, B.

Biochem. J. 309, 773-779, 1995

A:Title: Sequence variation of a novel heptahelical leucocyte receptor through alternative splicing

A:Reference number: S56162; MUID:95366951

A:Accession: S56162

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-327 <BAR>

A:Cross-references: EMBL:668829; NID:g840783; PIDN:CAA48723.1; PID:g840784

C:Superfamily: vertebrate rhodopsin

Query Match 96.7%: Score 29; DB 2; Length 327;
Best Local Similarity 80.0%: Pred. No. 1.4e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
1111
Db 307 FPSMR 311

RESULT 10

JT0585

minor endoglucanase (EC 3.2.1.-) precursor - Erwinia chrysanthemi

C:Species: Erwinia chrysanthemi

C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 08-Oct-1999

C:Accession: JT0585; JT0586

R:Guiseppi, A.; Aymeric, J.L.; Caml, B.; Barras, F.; Creuzet, N.

Gene 106, 109-114, 1991

A:Title: Sequence analysis of the cellulase-encoding cell gene of Erwinia chrysanthem

A:Reference number: JT0585; MUID:92039050

A:Accession: JT0585

A:Molecule type: DNA

A:Residues: 1-332 <GUI>

A:Cross-references: GB:M74044; NID:g148391; PIDN:AAA24818.1; PID:g148392

A:Experimental source: strain 3937

A:Accession: JT0586

A:Molecule type: protein

A:Residues: 24-33 <GUI>

C:Genetics:

A:Gene: cely

C:Keywords: glycosidase; hydrolase

F:1-23/Domain: signal sequence #status predicted <SIG>

F:24-332/Product: minor endoglucanase #status predicted <MIN>

Query Match 96.7%: Score 29; DB 2; Length 332;
Best Local Similarity 80.0%: Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
1111
Db 178 FPSMR 182

RESULT 11

G02913

sperm CD46 - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 09-Jun-2000

C:Accession: G02913

R:Hara, T.

submitted to GenBank, March 1996

A:Reference number: H01942

A:Accession: G02913

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-349 <HAR>

A:Cross-references: GB:D84105; NID:g1256700; PIDN:BAI12224.1; PID:g1256701

C:Superfamily: herpesvirus complement control protein; Complement factor H repeat hom

F:35-94/Domain: complement factor H repeat homology <FH01>

F:99-157/Domain: complement factor H repeat homology <FH02>

F:162-223/Domain: complement factor H repeat homology <FH03>

F:228-283/Domain: complement factor H repeat homology <FH04>

Query Match 96.7%: Score 29; DB 2; Length 349;
Best Local Similarity 80.0%: Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
1111
Db 11 FPSMR 15

RESULT 12
140696

endoglucanase - Cellulomonas uda
 C:Species: Cellulomonas uda
 C>Date: 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 05-Nov-1999
 C:Accession: I40696
 R:Nakamura, K.; Misawa, N.; Kitamura, K.
 J. Biotechnol. 4, 247-254, 1986
 A:Title: Sequence of a cellulase gene of Cellulomonas uda CB4.
 A:Reference number: I40696
 A:Accession: I40696
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-359 <RES>
 A:Cross-references: GB:M36503; NID:g144430; PIDN:AAA23090.1; PID:g144431

Query Match 96.7% Score 29; DB 2; Length 359;
 Best Local Similarity 80.0% Pred. No. 1.5e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 11 11
 DB 178 FPMR 182

RESULT 13
 157998
 membrane cofactor protein - human
 C:Species: Homo sapiens (man)
 C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 19-May-2000
 C:Accession: I57998
 R:Cervoni, F.; Fenichel, P.; Akhoundi, C.; Hsi, B.L.; Rossi, B.
 Mol. Reprod. Dev. 34, 107-113, 1993
 A:Title: Characterization of a cDNA clone coding for human testis membrane cofactor protein
 A:Reference number: I57998; MUID:93119658
 A:Accession: I57998
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-369 <RES>
 A:Cross-references: GB:S51940; NID:g262937; PIDN:AA824802.1; PID:g262938
 C:Superfamily: herpesvirus complement control protein; complement factor H repeat homolog
 F:35-94/Domain: complement factor H repeat homology <FH01>
 F:99-157/Domain: complement factor H repeat homology <FH02>
 F:162-223/Domain: complement factor H repeat homology <FH03>
 F:228-283/Domain: complement factor H repeat homology <FH04>

Query Match 96.7% Score 29; DB 2; Length 369;
 Best Local Similarity 80.0% Pred. No. 1.5e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 11 11
 DB 11 FPSMR 15

RESULT 14
 S26667
 G protein-coupled receptor BLR1 - human
 C:Species: Homo sapiens (man)
 C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jul-2000
 C:Accession: S26667
 R:Dobner, T.; Wolf, I.; Emrich, T.; Lipp, M.
 Eur. J. Immunol. 22, 2795-2799, 1992
 A:Title: Differential expression of a novel G protein-coupled receptor from
 A:Reference number: S26667; MUID:93049615
 A:Accession: S26667
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-372 <DOB>
 A:Cross-references: EMBL:X68149; NID:g29459; PIDN:CAA48252.1; PID:g29460
 C:Genetics:
 A:Gene: GDB:BLR1
 A:Cross-references: GDB:136235; OMIM:601613

A:Map position: 15q26.1-15q26.1
 C:Superfamily: vertebrate rhodopsin
 C:Keywords: G protein-coupled receptor; transmembrane protein

Query Match 96.7% Score 29; DB 2; Length 372;
 Best Local Similarity 80.0% Pred. No. 1.5e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 11 11
 DB 352 FPSMR 356

RESULT 15
 154479
 membrane cofactor protein precursor, splice form pm5.1 - human
 N:Alternate names: lymphocyte surface glycoprotein CD46
 C:Species: Homo sapiens (man)
 C>Date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 31-Mar-2000
 C:Accession: I54479
 R:Parcell, D.F.; Russell, S.M.; Deacon, N.J.; Brown, M.A.; Hooker, D.J.; McKenzie, I.
 Immunogenetics 33, 335-344, 1991
 A:Title: Alternatively spliced RNAs encode several isoforms of CD46 (MCP), a regulatory
 A:Reference number: I54479; MUID:91267562
 A:Accession: I54479
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-377 <RES>
 A:Cross-references: GB:M58050; NID:g180136; PIDN:AAA62833.1; PID:g180137
 C:Genetics:
 A:Gene: GDB:MCP
 A:Cross-references: GDB:120169; OMIM:120920
 A:Map position: 1q32-1q32
 C:Function:
 A:Description: for the factor I-mediated cleavage of the complement convertases
 C:Superfamily: herpesvirus complement control protein; complement factor H repeat homolog
 C:Keywords: alternative splicing; glycoprotein; transmembrane protein
 F:35-94/Domain: complement factor H repeat homology <FH01>
 F:99-157/Domain: complement factor H repeat homology <FH02>
 F:162-223/Domain: complement factor H repeat homology <FH03>
 F:228-283/Domain: complement factor H repeat homology <FH04>
 F:329-351/Domain: transmembrane #status predicted <TM>

Query Match 96.7% Score 29; DB 2; Length 377;
 Best Local Similarity 80.0% Pred. No. 1.6e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 11 11
 DB 11 FPSMR 15

Search completed: February 27, 2002, 11:45:23
 Job time: 702 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: February 27, 2002, 11:42:48 : Search time 78.39 Seconds
(without alignments)
2.339 Million cell updates/sec

Title: US-09-446-109a-15
Perfect score: 30
Sequence: 1 FPXMR 5

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_39.4

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	96.7	73	1 NXLI_OPHHA	P01387 ophiophagus
2	29	96.7	165	1 DSBV_VIBAL	O56578 vibrato algi
3	29	96.7	259	1 E2AA_ECOLI	P13810 escherichia
4	29	96.7	332	1 GUNY_ERWCH	P27032 erwina chr
5	29	96.7	359	1 GUNY_ERWCH	P18336 cellulomona
6	29	96.7	372	1 CCR5_HUMAN	P32302 homo sapien
7	29	96.7	377	1 MCP_HUMAN	P15329 homo sapien
8	29	96.7	420	1 2229_HUMAN	O9uj47 homo sapien
9	28	93.3	143	1 HS16_CABEL	P06581 caenorhabdi
10	28	93.3	143	1 HS16_CABEL	P02513 caenorhabdi
11	28	93.3	257	1 HNL_MANES	P52705 manihot esc
12	28	93.3	271	1 PABC_VIBHA	O56693 vibrato harv
13	28	93.3	271	1 PPKK_BACST	P58055 bacillus st
14	28	93.3	286	1 PAL_KLEPN	P37446 klebsiella
15	28	93.3	289	1 PAL_ECOLI	P00631 escherichia
16	28	93.3	289	1 PAL_PROVU	P37447 proteus vul
17	28	93.3	289	1 PAL_SALTY	P37442 salmonella
18	28	93.3	298	1 CC28_YEAST	P00546 saccharomyc
19	28	93.3	332	1 CC28_YEAST	P34556 caenorhabdi
20	28	93.3	374	1 CCR5_MOUSE	O04663 mus musculu
21	28	93.3	374	1 CCR5_MOUSE	O04663 mus musculu
22	28	93.3	430	1 KDTA_CHLMU	O9pk15 chlamydia m
23	28	93.3	431	1 KDTA_CHLTR	O57440 chlamydia t
24	28	93.3	433	1 ANM2_HUMAN	P55345 homo sapien
25	28	93.3	437	1 KDTA_CHLPN	O46222 chlamydia p
26	28	93.3	448	1 V347_CHLPN	O92836 chlamydia p
27	28	93.3	459	1 RBL2_THIDE	O60028 thiodacillu
28	28	93.3	463	1 RBL3_HYDMR	O59462 hydocogenov
29	28	93.3	470	1 LIP2_MYOCO	O64424 myocastor c
30	28	93.3	482	1 CAT2_ONCVC	O27710 onchocerca
31	28	93.3	490	1 CE05_ECOLI	O47500 escherichia
32	28	93.3	490	1 CE10_ECOLI	O47125 escherichia
33	28	93.3	492	1 CAT2_CUCPE	P48351 cucurbita p

34	28	93.3	492	1 CAT3_CUCPE	P48352 cucurbita p
35	28	93.3	548	1 CEAK_ECOLI	O47502 escherichia
36	28	93.3	766	1 EVA_DROME	O05201 drosophila
37	28	93.3	902	1 FTDH_RAT	P28037 rattus norv
38	28	93.3	973	1 TRP5_HUMAN	O9u162 homo sapien
39	28	93.3	974	1 TRP4_MOUSE	O9uq95 mus musculu
40	28	93.3	974	1 TRP5_RABIT	O62852 oryctolagus
41	28	93.3	975	1 TRP5_MOUSE	O9uq429 mus musculu
42	28	93.3	977	1 TRP4_HUMAN	O9uq44 homo sapien
43	28	93.3	977	1 TRP4_RAT	O35119 rattus norv
44	28	93.3	981	1 TRP4_BOVIN	P79100 bos taurus
45	28	93.3	1056	1 DPOL_ADE02	P03261 human adeno

ALIGNMENTS

RESULT	ID	Sequence	Standard	PRT	73 AA
AC	P01387	21-JUL-1986 (Rel. 01, Last sequence update)			
UT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	LONG NEUROTOXIN 1 (NEUROTOXIN A)				
OS	Ophiophagus hannah (King cobra)				
OC	Eukaryota: Metazoa: Chordata: Vertebrata: Euteleostomi: Lepidosauria: Squamata: Scleroglossa: Serpentes: Colubridae: Elapidae: Elapinae: Ophiophagus				
OX	NCBI_TaxID=6655				
FN	(1)				
RP	SEQUENCE				
IC	TISSUE=Venom				
RX	MEDLINE=73231298; PubMed=4198767				
KA	Joubert F.J.				
RT	"Snake venom toxins the amino acid sequences of two toxins from Ophiophagus hannah (King cobra) venom."				
RL	Biochim. Biophys. Acta 317:85-98(1973).				
CC	-1- SUBCELLULAR LOCATION: SECRETED.				
CC	-1- TISSUE SPECIFICITY: EXPRESSED BY VENOM GLAND.				
CC	-1- MISCELLANEOUS: LD(50) IS 0.3 MG/KG BY SUBCUTANEOUS INJECTION.				
CC	-1- SIMILARITY: BELONGS TO THE SNAKE TOXIN FAMILY.				
DR	HSSP: P01386; ITXB.				
DR	Interpro: IPR003571; Snake_toxin.				
DR	Pfam: PF00087; toxin.1.				
DR	ProDom: PD000206; Snake_toxin.1.				
DR	PROSITE: PS00272; SNAKE_TOXIN; 1.				
KW	Venom; Neurotoxin; Multigene family.				
FT	DISULFID 3 21				
FT	DISULFID 14 42				
FT	DISULFID 27 31				
FT	DISULFID 46 57				
FT	DISULFID 58 63				
FT	SEQUENCE 73 AA; 8106 MW; IAC17E91E16C54F7 CRC64;				
QY	1 FPXMR 5				
DB	66 FFWMR 70				
RESULT	2				
DSBB_VIBAL	STANDARD				
ID	DSBB_VIBAL				
AC	O56578				
DT	01-NOV-1997 (Rel. 35, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				

Query Match 96.7% Score 29; DB 1; Length 73;

Best Local Similarity 80.0% Pred. No. 17;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

DE DISULFIDE BOND FORMATION PROTEIN B (DISULFIDE OXIDOREDUCTASE)
DE (FRAGMENT).
GN DSB.
OS Vibrio alginolyticus.
OC Bacteria: Proteobacteria; gamma subdivision; Vibrionaceae; Vibrrio.
OX NCBI_TaxID=663;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B138-2;
RA Nakamura T., Enomoto H., to the EMBL/GenBank/DBJ databases.
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR DISULFIDE BOND FORMATION IN SOME
CC PERIPLASMIC PROTEINS. ACTS BY OXIDIZING THE DSB PROTEIN (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE DSB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D83728; AAA12087.1; -
DR InterPro: IPR003752; Dsdb.
DR Pfam: PF02600; Dsdb. 1.
KM Oxidoreductase; Redox-active center; Electron transport; Chaperone;
KW Transmembrane; Inner membrane.
FT DOMAIN 3 16 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 17 33 POTENTIAL.
FT DOMAIN 34 51 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 52 67 POTENTIAL.
FT DOMAIN 68 74 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 75 92 POTENTIAL.
FT DOMAIN 93 147 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 148 165 POTENTIAL.
FT DISULFID 43 46 REDOX-ACTIVE (BY SIMILARITY).
FT DISULFID 107 133 REDOX-ACTIVE (BY SIMILARITY).
FT NON_TER 165 165
SQ SEQUENCE 165 AA; 18745 MW; D43F96DF5AA3272D CRC64;

Query Match 96.7%; Score 29; DB 1; Length 165;
Best Local Similarity 80.0%; Pred. No. 37;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
DB 113 FPMWR 117

RESULT 3
E2AA_ECOLI STANDARD; PRT; 259 AA.
ID P13810;
AC 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HEAT-LABILE ENTEROTOXIN IIA, A CHAIN PRECURSOR (LT-IIA).
OS Escherichia coli.
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=88032841; PubMed=2822667;
RA Pickett C.L., Weinstein D.L., Holmes R.K.;
RT "Genetics of type IIA heat-labile enterotoxin of Escherichia coli:
RT operon fusions, nucleotide sequence, and hybridization studies.";
RL J. Bacteriol. 169:5180-5187(1987).

```

```

CC -1- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY
CC THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.
CC -1- SUBUNIT: HETEROHETEROMER OF ONE A CHAIN AND OF FIVE B CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M17894; AAA24093.1; -
DR PIR: A29831; A29831.
DR HSSP: PA3528; ITII.
DR InterPro: IPR001144; Enterotoxin_A.
DR Pfam: PF01375; Enterotoxin_A. 1.
KW Enterotoxin; Signal.
FT SIGNAL 1 18
FT CHAIN 19 259 HEAT-LABILE ENTEROTOXIN IIA, A CHAIN.
FT DISULFID 203 215 BY SIMILARITY.
FT ACT_SITE 128 128
SQ SEQUENCE 259 AA; 29242 MW; 996F311A32CABEAA CRC64;

Query Match 96.7%; Score 29; DB 1; Length 259;
Best Local Similarity 80.0%; Pred. No. 56;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
DB 187 FPMWR 191

RESULT 4
GUNY_ERMCH STANDARD; PRT; 332 AA.
ID AC P27032;
AC 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE MINOR ENDOGLUCANASE Y PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE
DE Y) (CELLULOSE Y) (EGY).
CN CELY
OS Erwiniia chrysanthemi.
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=556;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-33.
RC STRAIN=3937;
RC MEDLINE=92039050; PubMed=1937031;
RA Guiseppi A., Aymeric J.-L., Caml D., Barras F., Creuzet N.;
RT "Sequence analysis of the cellulase-encoding cely gene of Erwiniia
RT chrysanthemi: a possible case of interspecies gene transfer.";
RL Gene 106:109-114(1991).
CC -1- FUNCTION: REPRESENTS ONLY 3% OF THE GLOBAL CELLULOSE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY D (FAMILY 8 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M74044; AAA24818.1; -
DR PIR: J70585; J70585.

```

DR InterPro: IPR002037; Glyco_hydro-8.
 DR Pfam: PF01270; Glyco_hydro-8; 1.
 DR PRINTS: PR00735; GLYDRLASE8.
 DR PROSITE: PS00812; GLYCOSYL_HYDROL_F8; 1.
 KM Cellulose degradation; Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 23
 FT CHAIN 24 332 MINOR ENDOGLUCANASE Y.
 FT ACT_SITE 53 53 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 110 110 NUCLEOPHILE (POTENTIAL).
 SQ SEQUENCE 332 AA: 37592 MW: 6EA760A2B8227079 CRC64:

Query Match: 96.7%; Score 29; DB 1; Length 332;
 Best Local Similarity 80.0%; Pred. No. 71;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
 DB 178 FPAWR 182

RESULT 5
 GUN_CELUD STANDARD; PRT: 359 AA.
 ID GUN_CELUD
 AC P18336;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
 DE (CELLULOSE).
 OS Cellulomonas uda.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Micrococcales; Cellulomonadaceae; Cellulomonas.
 OX NCBI_TaxID=1714;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-37.
 RC STRAIN=CB4;
 RA Nakamura K., Misawa N., Kitamura K.;
 RT J. Biotechnol. 4:247-254(1986).
 RL *Sequence of a cellulase gene of Cellulomonas uda CB4.*.
 CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
 CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
 CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
 CC (2) EXOCELLULOBIODHOLASES THAT CUT THE DISACCHARIDE CELLULOSE
 CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
 CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER
 CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY D (FAMILY 8 OF GLYCOSYL
 CC HYDROLASES).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to [license@isb-sid.ch](mailto:license@isb-sib.ch)).
 CC
 CC EMBL: M36503; AAA23090.1;
 DR InterPro: IPR002037; Glyco_hydro-8.
 DR Pfam: PF01270; Glyco_hydro-8; 1.
 DR PROSITE: PS00812; GLYCOSYL_HYDROL_F8; 1.
 KM Cellulose degradation; Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 23
 FT CHAIN 24 359 ENDOGLUCANASE.
 FT ACT_SITE 53 53 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 110 110 NUCLEOPHILE (POTENTIAL).
 SQ SEQUENCE 359 AA: 40690 MW: 0445D7571B683148 CRC64:

Query Match 96.7%; Score 29; DB 1; Length 359;

Best Local Similarity 80.0%; Pred. No. 76;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
 DB 178 FPAWR 182

RESULT 6
 CCR5_HUMAN STANDARD; PRT: 372 AA.
 ID CCR5_HUMAN
 AC P32302; Q14811;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE C-X-C CHEMOKINE RECEPTOR TYPE 5 (CXCR-5) (CXCR-5) (BURKITT'S LYMPHOMA
 DE RECEPTOR 1) (MONOCYTE-DERIVED RECEPTOR 15) (MDR15).
 GN BLR1 OR CXCR5.
 OS Homo sapiens (human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG FORM).
 RC TISSUE=Lymphocytes;
 KC MEDLINE=93049615; PubMed=1425907;
 KA Dobner T., Wolf I., Emrich T., Lipp M.;
 RT "Differentiation-specific expression of a novel G protein-coupled
 RT receptor from Burkitt's lymphoma."
 RL Eur. J. Immunol. 22:2795-2799(1992).
 EN [2]
 RP SEQUENCE FROM N.A. (SHORT FORM).
 RC TISSUE=Blood;
 KC MEDLINE=95366951; PubMed=7639692;
 KA Barelle L., Loetscher M., Tobler A., Baggiolini M., Moser B.;
 RT "Sequence variation of a novel heptahelical leucocyte receptor
 RT through alternative transcript formation."
 RL Biochem. J. 309:773-779(1995).
 PN [3]
 RP LIGAND BINDING.
 KC MEDLINE=98130629; PubMed=9463416;
 RA Legler D.F., Loetscher M., Stuber Roos R., Clark-Iewis I.,
 RA Baggiolini M., Moser B.;
 RT "B cell-attracting chemokine 1, a human CXC chemokine expressed in
 RT lymphoid tissues, selectively attracts B lymphocytes via
 RT BLR1/CXCR5."
 RL J. Exp. Med. 187:655-660(1998).
 CC -1- FUNCTION: CYTOKINE RECEPTOR THAT BINDS TO BLC. BLR1 EXERTS
 CC POSSIBLY A REGULATORY FUNCTION IN BURKITT LYMPHOMA (BL).
 CC LYMPHOMAGENESIS AND/OR B-CELL DIFFERENTIATION. IT IS A POTENTIAL
 CC CANDIDATE FOR CELL-CELL INTERACTION, AND ACTIVATION OF MATURE B-
 CC LYMPHOCYTES IN LYMPHATIC TISSUES.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSION IN MATURE B-CELLS AND BURKITT
 CC LYMPHOMA CELLS.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X68149; CAA48252.1;
 DR EMBL: X68829; CAA48723.1;
 DR PIR: S26667; S26667.
 DR GCRDB: GCR_0453;
 DR GCRDB: GCR_2072;
 DR GCRDB: GCR_2612;


```

FT DOMAIN 35 328 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 329 351 POTENTIAL.
FT DOMAIN 352 377 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 35 95 SUSHI 1.
FT DOMAIN 98 158 SUSHI 2.
FT DOMAIN 161 224 SUSHI 3.
FT DOMAIN 227 284 SUSHI 4.
FT DOMAIN 287 311 SER/THR-RICH.
FT DISULFID 35 80 BY SIMILARITY.
FT DISULFID 64 94 BY SIMILARITY.
FT DISULFID 99 141 BY SIMILARITY.
FT DISULFID 127 157 BY SIMILARITY.
FT DISULFID 162 210 BY SIMILARITY.
FT DISULFID 191 223 BY SIMILARITY.
FT DISULFID 228 270 BY SIMILARITY.
FT DISULFID 256 283 BY SIMILARITY.
FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 273 273 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPIC 286 300 MISSING (IN A SECOND FORM).
FT VARSPIC 340 352 MISSING (IN ISOFORM M).
FT VARSPIC 353 361 YLORRRKKG -> DIFRGRRKKQKVELMPLRLNPLQ
FT VARSPIC 362 377 QSRRAE (IN ISOFORM N).
FT VARSPIC 362 377 TYLTDETHREVKFTSL -> KADCAEYATVQKSTTPAQ
FT VARSPIC 362 377 RG (IN ISOFORM B/D/F/H/J/L).
FT VARSPIC 340 361 MISSING (IN ISOFORM M AND ISOFORM N).
FT VARSPIC 339 339 MISSING (IN ISOFORM N).
FT VARSPIC 339 339 I -> IGKQVELMPLRLNPLQOSREA (IN
SQ SEQUENCE 377 AA: 42247 MW: 2CA6F61752570B57 CRC64:
  Query Match 96.7%: Score 29; DB 1: Length 377;
  Best Local Similarity 80.0%: Pred. No. 80;
  Matches 4: Conservative 0; Mismatches 1; Indels 0; Gaps 0;
  QY 1 FPXWR 5
  DB 11 FPSWR 15
  RESULT 8
  2229_HUMAN STANDARD: PRT: 420 AA.
  AC Q9UJW7:
  DT 20-AUG-2001 (Rel. 40, Created)
  DT 20-AUG-2001 (Rel. 40, Last sequence update)
  DT 20-AUG-2001 (Rel. 40, Last annotation update)
  DE ZINC FINGER PROTEIN 229 (FRAGMENT).
  CN ZNF229.
  OS Homo sapiens (Human).
  OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
  OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
  OX NCBI_TaxID=9606;
  RN [1]
  RP SEQUENCE FROM N.A.
  RA Shannon M., Branscomb E., Hauser L., Gordon L., Ashworth L.,
  RA Stubbs L.;
  RT "Differential expansion of homologous zinc-finger gene families in
  RT human chromosome 19q13.2 and mouse chromosome 7."
  RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
  RP SEQUENCE FROM N.A.
  RA Kodyanov V., Ge Y., Krummel G.K., Kvistad E., Grable L.,
  RA Severin J., Gordon L., Shannon M., Brower A., Olsen A.S., Smith L.M.;
  RT "Sequence analysis of a 1mb region in 19q13.2 containing a zinc finger
  RT gene cluster."
  RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
  CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
  CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
  CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
  CC -1- SIMILARITY: CONTAINS 1 KRAE DOMAIN.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF192979; AAF07964.1; -
DR EMBL: AC084239; AAG23970.1; -
DR InterPro: IPR001909; KRAE.
DR InterPro: IPR000822; ZnF-C2H2.
DR Pfam: PF01352; KRAE.1.
DR Pfam: PF00096; zf-C2H2; 2.
DR PRINTS: PR00048; ZINC_FINGER.
DR SMART: SM00349; KRAE.1.
DR SMART: SM00355; ZnF-C2H2; 2.
DR PROSITE: PS00805; KRAE.1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 2.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 3.
DR Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 34 108 KRAE.
FT DOMAIN 349 >420 ZINC_FINGERS.
FT ZN_FING 349 371 C2H2-TYPE.
FT ZN_FING 377 399 C2H2-TYPE.
FT ZN_FING 405 >420 C2H2-TYPE.
FT NON_TER 420 420
SQ SEQUENCE 420 AA: 48022 MW: FA4138BA44A2A14 CRC64:
  Query Match 96.7%: Score 29; DB 1: Length 420;
  Best Local Similarity 80.0%: Pred. No. 89;
  Matches 4: Conservative 0; Mismatches 1; Indels 0; Gaps 0;
  QY 1 FPXWR 5
  DB 173 FPAWR 177
  RESULT 9
  HS16_CAEEL STANDARD: PRT: 143 AA.
  AC P06581; P02514;
  DT 01-JAN-1988 (Rel. 06, Created)
  DT 01-JAN-1988 (Rel. 06, Last sequence update)
  DT 01-FEB-1996 (Rel. 33, Last annotation update)
  DE HEAT SHOCK PROTEIN HSP16-41.
  CN HSP16-41.
  OS Caenorhabditis elegans.
  OC Eukaryota; Metazoa; Nematoda; Chordata; Rhabdita; Rhabditidae;
  OC Rhabditidae; Pelodermatidae; Caenorhabditis.
  OX NCBI_TaxID=6239;
  RN [1]
  RP SEQUENCE FROM N.A.
  RA MEDLINE:86304344; PubMed:3017958;
  RA Jones D., Russnak R.H., Kay R.J., Candido E.P.M.;
  RT "Structure, expression, and evolution of a heat shock gene locus in
  RT Caenorhabditis elegans that is flanked by repetitive elements."
  RL J. Biol. Chem. 261:12006-12015(1986).
  RP SEQUENCE OF 47-143 FROM N.A.
  RX MEDLINE:83220736; PubMed:6190129;
  RA Russnak R.H., Jones D., Candido E.P.M.;
  RT "Cloning and analysis of cDNA sequences coding for two 16 kilodalton
  RT heat shock proteins (hsps) in Caenorhabditis elegans: homology with
  RT the small hsps of Drosophila."
  RL Nucleic Acids Res. 11:3187-3205(1983).
  CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
  CC FAMILY.
  CC -----
  CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M14334: AAA28070.1: ALT-SEQ.
 DR EMBL: X01577: CAA25732.1: -
 DR PIR: A38884: HHKMA1.
 DR PIR: A25199: A25199.
 DR InterPro: IPR002068: Crystallin_HSP20.
 DR Pfam: PF00011: HSP20: 1.
 DR PROSITE: PS01031: HSP20: 1.
 DR Heat shock: Multigene family.
 DR SEQUENCE 143 AA: 16252 MW: C1D0F59D26E36C24 CRC64;
 SO
 Query Match 93.3% Score 28: DB 1: Length 143:
 Best Local Similarity 80.0% Pred. No. 50:
 Matches 4: Conservative 0: Mismatches 1: Indels 0: Gaps 0:
 Oy 1 FPXMR 5
 11 11
 27 FPYMR 31
 Db
 RESULT 10
 ID HS17-CAEEL STANDARD: PRT: 143 AA.
 AC P02513:
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE HEAT SHOCK PROTEIN HSP16-48.
 CN HSP16-48.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85295957; PubMed=4033652;
 RA Russnak R.H., Candido E.P.M.:
 RT "Locs encoding a family of small heat shock genes in Caenorhabditis
 RT elegans: two genes duplicated to form a 3.8-kilobase inverted
 RT repeat".
 RT Mol. Cell. Biol. 5:1268-1276(1985).
 RL [2]
 RN
 RP SEQUENCE OF 9-143 FROM N.A.
 RX MEDLINE=83220736; PubMed=6190129;
 RA Russnak R.H., Jones D., Candido E.P.M.:
 RT "Cloning and analysis of cDNA sequences coding for two 16 kilodalton
 RT heat shock proteins (hsps) in Caenorhabditis elegans: homology with
 RT the small hsps of Drosophila".
 RL Nucleic Acids Res. 11:3187-3205(1983).
 CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: K03273: AAA28069.1: -
 DR EMBL: X01576: CAA25731.1: -
 DR EMBL: K01863: AAA28064.1: -
 DR PIR: A02016: HHKMA8.
 DR PIR: A24289: A24289.
 DR InterPro: IPR002068: Crystallin_HSP20.

DR Pfam: PF00011: HSP20: 1.
 DR PROSITE: PS01031: HSP20: 1.
 DR Heat shock: Multigene family.
 DR SEQUENCE 143 AA: 16299 MW: 0D5596DFEE5B3318 CRC64;
 SO
 Query Match 93.3% Score 28: DB 1: Length 143:
 Best Local Similarity 80.0% Pred. No. 50:
 Matches 4: Conservative 0: Mismatches 1: Indels 0: Gaps 0:
 Oy 1 FPXMR 5
 11 11
 27 FPYMR 31
 Db
 RESULT 11
 ID HNL-MANES STANDARD: PRT: 257 AA.
 AC P52705:
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE (S)-ACETONE-CYANOHYDRIN LYASE (EC 4.1.2.39) ((S)-HYDROXYNITRILE LYASE)
 DE ((S)-HYDROXYNITRILASE) (OXYNITRILASE).
 CN HNL.
 OS Manihot esculenta (Cassava) (Manioc).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids 1; Malpighiales; Euphorbiaceae; Manihot.
 OX NCBI_TaxID=3583;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-36 AND 168-191.
 HC TISSUE=Cotyledon;
 RX MEDLINE=94263231; PubMed=8203915;
 EA Hughes J., Carvalho F.J.P.D.C., Hughes M.A.:
 RT "Purification, characterization, and cloning of alpha-hydroxynitrile
 RT lyase from cassava (Manihot esculenta Gratz).".
 RL Arch. Biochem. Biophys. 311:496-502(1994).
 CC -1- FUNCTION: INVOLVED IN CYANOGENESIS. THE RELEASE OF HCN FROM
 CC INJURED TISSUES. DECOMPOSES A VARIETIES OF (R) OR (S) CYANOHYDRINS
 CC INTO HCN AND THE CORRESPONDING ALDEHYDES AND KETONES. THE NATURAL
 CC SUBSTRATE OF THIS ENZYME IS (S)-ACETONE CYANOHYDRIN.
 CC -1- CATALYTIC ACTIVITY: (S)-2-HYDROXYISOBUTYRONITRILE = CYANIDE +
 CC ACETONE.
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- SIMILARITY: DISTANT RELATIONSHIP WITH TYPE-B
 CC CARBOXYL-ESTERASE/LIPASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z29091: CAA02334.1: -
 DR HSSP: P52704: LYAS.
 DR InterPro: IPR000073: Abhydrolase.
 DR InterPro: IPR000379: Est_Lip_thioest_actsite.
 DR Pfam: PF00561: abhydrolase; 1.
 KW Lyase.
 FT INIT_MET 0 0
 FT ACT_SITE 79 79 BY SIMILARITY.
 FT ACT_SITE 207 207 BY SIMILARITY.
 FT ACT_SITE 235 235 BY SIMILARITY.
 FT ACT_SITE 235 235 BY SIMILARITY.
 SO SEQUENCE 257 AA: 29240 MW: 98B3E160ACB338C5 CRC64;
 Query Match 93.3% Score 28: DB 1: Length 257:
 Best Local Similarity 80.0% Pred. No. 87:
 Matches 4: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

OY 1 FPXWR 5
11 11
Db 124 FPDWR 128

RESULT 12

PABC_VIBHA STANDARD: PRT: 271 AA.
ID PABC_VIBHA
AC 056693;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 40, Last annotation update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE 4-AMINO-4-DEOXYCHORISMATE LYASE (EC 4.1.3.1) (ADC LYASE) (ADCL).
GN PABC.
OS Vibrio harveyi.
OC Bacteria: Proteobacteria: gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=669;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=8392;
RX MEDLINE=96134997; PubMed=8550484;
RA Shen Z., Byers D.M.;
RT and fabf genes involved in fatty acid biosynthesis.*;
RL J. Bacteriol. 178:571-573(1996).
CC -1- FUNCTION: CONVERTS 4-AMINO-4-DEOXYCHORISMATE INTO 4-AMINOENZOATE (PABA) AND PYRUVATE.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -1- PATHWAY: SECOND STEP IN FOLATE BIOSYNTHESIS PATHWAY.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT AMINOTRANSFERASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: U39441; AAC43592.1;
DR Interpro: IPR001544; AminoTran_4.
DR Pfam: PFO1063; aminoTran_4; 1.
DR ProDom: PD001961; AminoTran_4; 1.
DR PROSITE: PS00770; AA_TRANSFER_CLASS_4; 1.
KW Lyase: Pyridoxal phosphate; Folate biosynthesis.
FT BINDING 140 140 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 271 AA: 29958 MW: 3AEBEB9F398D1CD CRC64;

Query Match 93.3%; Score 28; DB 1; Length 271;
Best Local Similarity 80.0%; Pred. No. 91;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
11 11
Db 57 FPDWR 61

RESULT 13

PNK_BACST STANDARD: PRT: 271 AA.
ID PNK_BACST
AC P58055;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE INORGANIC POLYPHOSPHATE/ATP-NAD KINASE (EC 2.7.1.23)
DE (POLY(P)/ATP NAD KINASE).
GN PNK OR NADK.
OS Bacillus steaerothermophilus.
OC Bacteria: Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Geobacillus.

OX NCBI_TaxID=1422;
PN (1)
RP SEQUENCE FROM N.A.
RA Ohshima T., Sakuraba H.;
RT "Inorganic polyphosphate/ATP-NAD kinase of Bacillus
steaerothermophilus.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CATALYZES THE PHOSPHORYLATION OF NAD TO NADP. UTILIZES
ATP AND OTHER NUCLEOSIDE TRIPHOSPHATES AS WELL AS INORGANIC
POLYPHOSPHATE AS A SOURCE OF PHOSPHORUS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + NAD(+) -> ADP + NADP(+) +
H₂O.
CC -1- COFACTOR: REQUIRES DIVALENT METAL IONS FOR ACTIVITY (BY
SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NAD KINASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: AB059961; BAB32727.1;
DR Transferase: Kinase: NADP.
KW Nucleoside triphosphatase.
SQ SEQUENCE 271 AA: 30517 MW: 9E0FB427BC76662A CRC64;

Query Match 93.3%; Score 28; DB 1; Length 271;
Best Local Similarity 80.0%; Pred. No. 91;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
11 11
Db 258 FPDWR 262

RESULT 14

PAL_KLEPN STANDARD: PRT: 286 AA.
ID PAL_KLEPN
AC P37446;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PHOSPHOLIPASE A1 PRECURSOR (EC 3.1.1.32) (DETERGENT-RESISTANT
DE PHOSPHOLIPASE A) (DR-PHOSPHOLIPASE A) (PHOSPHATIDYLCHOLINE 1-
DE ACYLHYDROLASE) (OUTER MEMBRANE PHOSPHOLIPASE A) (OM PLA).
GN PLDA.
OS Klebsiella pneumoniae.
OC Bacteria: Proteobacteria: gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=573.
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=94131966; PubMed=8300539;
RA Brok R.G.P.M., Brinkman F., van Bortel R., Bekkers A.C.A.P.,
Verheij H.M., Tomassen J.;
RT Molecular characterization of enterobacterial plda genes encoding
outer membrane phospholipase A.*;
RL J. Bacteriol. 176:861-870(1994).
CC -1- FUNCTION: HYDROLYSIS OF PHOSPHATIDYLCHOLINE WITH PHOSPHOLIPASE
A2 (EC 3.1.1.4) AND PHOSPHOLIPASE A1 (EC 3.1.1.32) ACTIVITIES.
CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H₂O -> 1-ACYLGLYCERO-
PHOSPHOCHOLINE + A FATTY ACID ANION.
CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H₂O -> 2-ACYLGLYCERO-
PHOSPHOCHOLINE + A FATTY ACID ANION.
CC -1- COFACTOR: REQUIRES CALCIUM IONS FOR ACTIVITY.
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE; ONE OF THE VERY FEW ENZYMES
LOCATED THERE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL: X76901: CAA54223.1: -
DR PIR: B36971: B36971.
DR PIR: S40129: S40129.
DR InterPro: IPR003187: PLA1.
DR Pfam: PF02253: PLA1: 1.
KM HydroLase: Lipid degradation: Outer membrane: Signal: Calcium.
FT SIGNAL: 21 20 BY SIMILARITY.
FT CHAIN: 21 286 PHOSPHOLIPASE A1.
FT ACT_SITE: 161 161 BY SIMILARITY.
SQ SEQUENCE 286 AA: 32544 MW: 3239F863085108A3 CRC64;

Query Match 93.38: Score 28: DB 1: Length 286;
Best Local Similarity 80.0%: Pred. No. 96;
Matches 4: Conservative 0: Mismatches 1: Indels 0: Gaps 0;

Oy 1 FPXMR 5
Db 92 PPLMR 96

RESULT 15
PAL_ECOLI STANDARD: PRT: 289 AA.
ID PAL_ECOLI
AC P00631:
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PHOSPHOLIPASE A1 PRECURSOR (EC 3.1.1.32) (DETERGENT-RESISTANT
DE PHOSPHOLIPASE A) (DR-PHOSPHOLIPASE A) (PHOSPHATIDYLCHOLINE 1-
DE ACYLHYDROLASE) (OUTER MEMBRANE PHOSPHOLIPASE A) (OM PLA).
GN PLDA OR B3821 OR 25342 OR ECS4751.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85157492: PubMed=6397464;
RA Homma H., Kobayashi T., Chiba N., Karasawa K., Mizushima H., Kudo I.,
RA Inoue K., Ikeda H., Sekiguchi M., Nojima S.;
RT "The DNA sequence encoding plda gene, the structural gene for
RT detergent-resistant phospholipase A of E. coli.";
RL J. Biochem. 96:1655-1664(1984).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=92358234: PubMed=1379743;
RA Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;
RT "Analysis of the Escherichia coli genome: DNA sequence of the region
RT from 84.5 to 86.5 minutes.";
RL Science 257:771-778(1992).
RN [3]
RP REVISION TO 14-15.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617: PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935: PubMed=11206551;

HA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackert J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
HA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
HA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231: PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
HA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hatiori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [6]
RP SEQUENCE OF 30-289 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=85003590: PubMed=6383820;
RA de Geus P., Verheij H.M., Riegman N.H., Hoekstra W.P.M., de Haas G.H.;
RT "The pro- and mature forms of the E. coli K-12 outer membrane
RT phospholipase A are identical.";
RL EMBO J. 3:1799-1802(1984).
RN [7]
RP SEQUENCE OF 174-289 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=87115164: PubMed=3027506;
RA Iino N., Nakayama K., Nakayama H.;
RT "The recG gene of Escherichia coli K12: primary structure and
RT evidence for SOS regulation.";
RL Mol. Gen. Genet. 205:298-304(1986).
RN [8]
RP NOTAGENESIS OF SER-172.
RX MEDLINE=94131966: PubMed=8300539;
RA Brok R.G.P.M., Brinkman E., van Boxtel R., Bekkers A.C.A.P.,
RA Verheij H.M., Tomassen J.;
RT "Molecular characterization of enterobacterial plda genes encoding
RT outer membrane phospholipase A.";
RL J. Bacteriol. 176:861-870(1994).
RN [9]
RP ACTIVE SITE SER-164, AND PARTIAL SEQUENCE.
RX MEDLINE=91249806: PubMed=2040286;
RA Horrevorts A.J.G., Verheij H.M., de Haas G.H.;
RT "Inactivation of Escherichia coli outer-membrane phospholipase A by
RT the affinity label hexadecanesulfonyl fluoride. Evidence for an
RT active-site serine.";
RL Eur. J. Biochem. 198:247-253(1991).
RN [10]
RP FUNCTION: HYDROLYSIS OF PHOSPHATIDYLCHOLINE WITH PHOSPHOLIPASE
RC A2 (EC 3.1.1.4) AND PHOSPHOLIPASE A1 (EC 3.1.1.32) ACTIVITIES.
RX REQUIRED FOR EFFICIENT SECRETION OF BACTERIOTOXINS, SEEMS TO BE
RC DORMANT IN NORMAL GROWING CELLS.
CC CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 1-ACYLGLYCERO-
CC PHOSPHOCHOLINE + A FATTY ACID ANION.
CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 2-ACYLGLYCERO-
CC PHOSPHOCHOLINE + A FATTY ACID ANION.
CC -1- COFACTOR: REQUIRES CALCIUM IONS FOR ACTIVITY.
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE; ONE OF THE VERY FEW ENZYMES
CC LOCATED THERE.
CC -1- INDUCTION: BY MEMBRANE DAMAGING, FOR EXAMPLE, BY PHAGE-INDUCED
CC LYSIS OR TEMPERATURE SHOCK.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X02143: CAA26081.1; -
DR EMBL: M87049: AAC67617.1; -
DR EMBL: AE000458: AAC76824.1; -
DR EMBL: AE005613: AAC59017.1; -
DR EMBL: AP002567: BAB38174.1; -
DR EMBL: M30198: AAA24516.1; -
DR PIR: A00771: PSECA.
DR PIR: A22133: PSECA1.
DR PIR: S30711: S30711.
DR Ecogene: EG10738: plda.
DR InterPro: IPR003187: PLA1.
DR Pfam: PF02253: PLA1; 1.
KW Hydrolyase; Lipid degradation; Outer membrane; Signal; Calcium;
KW Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 289 PHOSPHOLIPASE A1.
FT ACT_SITE 164 164
FT MUTAGEN 172 172 S->F: INACTIVE PROTEIN.
FT CONFLICT 14 15 LP -> FA (IN REF. 2).
FT CONFLICT 30 33 DAPA -> MTRQ (IN REF. 6).
SQ SEQUENCE 289 AA: 33163 MW: A688AD32AA60F218 CRC64;

Query Match 93.3%; Score 28; DB 1; Length 289;
Best Local Similarity 80.0%; Pred. No. 97;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 FPXWR 5
DB 95 FPLMR 99

Search completed: February 27, 2002, 11:42:48
Job time: 547 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:50:14 ; Search time 281.76 Seconds
(without alignments)
2.596 Million cell updates/sec

Title: US-09-446-109a-15

Perfect score: 30

Sequence: 1 FPXMR 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_17: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phage: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	29	96.7	101 5 Q9V671	Q9V671 drosophila
2	29	96.7	113 5 Q39293	Q39293 brassica na
3	29	96.7	157 5 Q9XRV5	Q9XRV5 caenorhabd
4	29	96.7	158 4 Q9NNM3	Q9NNM3 homo sapien
5	29	96.7	158 4 Q9NNM2	Q9NNM2 homo sapien
6	29	96.7	229 8 Q37633	Q37633 thagolelis
7	29	96.7	245 2 Q9PDE1	Q9PDE1 xylella fas
8	29	96.7	273 10 Q9SUUV	Q9SUUV arabidopsis
9	29	96.7	292 2 Q9H783	Q9H783 pseudomonas
10	29	96.7	321 11 Q9D6L7	Q9D6L7 mus musculu
11	29	96.7	332 2 Q9APJ5	Q9APJ5 erwinia chr
12	29	96.7	333 2 Q9L3G9	Q9L3G9 erwinia rha
13	29	96.7	334 10 Q9FTM4	Q9FTM4 oryza sativ
14	29	96.7	349 4 Q15429	Q15429 homo sapien
15	29	96.7	417 10 Q9SAAL	Q9SAAL arabidopsis
16	29	96.7	466 2 Q9FDK4	Q9FDK4 zymomonas m
17	29	96.7	479 2 Q86742	Q86742 streptomyces
18	29	96.7	497 2 Q91702	Q91702 pseudomonas
19	29	96.7	501 5 Q16923	Q16923 caenorhabd

20	29	96.7	502 5 Q9GCU15	Q9GCU15 caenorhabd
21	29	96.7	555 1 Q29748	Q29748 archaeglob
22	29	96.7	581 4 Q9H801	Q9H801 homo sapien
23	29	96.7	612 4 Q9H8Y2	Q9H8Y2 homo sapien
24	29	96.7	621 4 Q9H9Y1	Q9H9Y1 homo sapien
25	29	96.7	625 2 Q66676	Q66676 aquilex aco
26	29	96.7	634 4 Q9NM19	Q9NM19 homo sapien
27	29	96.7	642 4 Q9NMX5	Q9NMX5 homo sapien
28	29	96.7	642 2 Q25396	Q25396 heliobacte
29	29	96.7	642 2 Q97LE3	Q97LE3 heliobacte
30	29	96.7	650 5 Q17866	Q17866 caenorhabd
31	29	96.7	735 2 Q9HU63	Q9HU63 pseudomonas
32	29	96.7	778 5 Q9U9K6	Q9U9K6 caenorhabd
33	29	96.7	811 4 Q9BMX2	Q9BMX2 homo sapien
34	29	96.7	1071 11 Q9D2K4	Q9D2K4 mus musculu
35	29	96.7	1108 13 Q9PMD0	Q9PMD0 tetraodon f
36	29	96.7	1117 5 Q9U9K7	Q9U9K7 caenorhabd
37	29	96.7	1451 2 Q45336	Q45336 bordetella
38	29	96.7	1451 2 Q45044	Q45044 bordetella
39	29	96.7	1464 2 Q9S5D5	Q9S5D5 bordetella
40	28	93.3	68 2 P73858	P73858 synecocyst
41	28	93.3	84 2 P74471	P74471 synecocyst
42	28	93.3	84 2 Q55568	Q55568 synecocyst
43	28	93.3	93 2 P73028	P73028 synecocyst
44	28	93.3	103 10 Q9M045	Q9M045 arabidopsis
45	28	93.3	110 6 Q9GLJ9	Q9GLJ9 canis famli

ALIGNMENTS

RESULT 1
Q9V671 PRELIMINARY: PRT: 101 AA.
AC Q9V671: 01-MAY-2000 (TREMBLrel, 13, Created)
DT 01-MAY-2000 (TREMBLrel, 13, Last sequence update)
UT 01-JUN-2001 (TREMBLrel, 17, Last annotation update)
DE CG13174 PROTEIN.
CN CG13174
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.C., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Neilson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burks K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harits M.,
RA Harits N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeigyan C.,
RA Jatalin M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA Lasro P., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maliet B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merulov G., Milshina N.V., Moberg C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA SVRskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin C.M., Venter J.C.;
 RA "The genome-sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AEO03823; AAF58560.1; .
 DR FlyBase: FBgn0033694; CG13174.
 SO SEQUENCE 101 AA: 11820 MW: D5C4D85B514CB89B CRC64;

Query Match 96.7%; Score 29; DB 5; Length 101;
 Best Local Similarity 80.0%; Pred. No. 1.2e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 DB 75 FPMWR 79

RESULT 2
 O39293 PRELIMINARY: PRT; 113 AA.

AC 039293; 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE ELONGATION FACTOR EF-1A (FRAGMENT).
 OS Brassica napus (rape).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Brassica.
 OX NCBL_Taxid=3708;

RP SEQUENCE FROM N.A.
 RC STRAIN=SAMOUEAL; TISSUE=ETIOLATED SEEDLINGS, (5-DAYS-OLD);
 RA Saez-Vasquez J.;
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO GTP-BINDING ELONGATION FACTOR FAMILY.
 DR EMBL: U21744; AAA86366.1; .
 DR Mendel; 15767; Brana; 1139; 15767.
 DR InterPro: IPR000795; GTP_EFTU.
 DR Pfam: PF00009; GTP_EFTU; 1.
 KM Elongation factor: GTP-binding; Protein biosynthesis.
 FT NON_TER
 SO SEQUENCE 113 AA: 12560 MW: 1497830707223948 CRC64;

Query Match 96.7%; Score 29; DB 10; Length 113;
 Best Local Similarity 80.0%; Pred. No. 1.4e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 DB 103 FPMWR 107

RESULT 3
 O9XTV5 PRELIMINARY: PRT; 157 AA.

AC 09XTV5; 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE T03D8.2 PROTEIN.

CN T03D8.2.
 OS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBL_Taxid=6239;

RP SEQUENCE FROM N.A.
 RA Mortimore B.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 RN (2)

RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sultson J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterson R.,
 RA Watson A., Weinstein L., Wilkinson-Sproat J., Woldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*.";
 RL Nature 368:32-38(1994).
 DR EMBL: 292838; CAB07406.1; .
 DR InterPro: IPR000230; Ribosomal_S12.
 DR Pfam: PF00164; Ribosomal_S12; 1.
 DR PRINTS: PR01034; RIBOSOMAL_S12.
 DR PRODOM: PD000576; Ribosomal_S12; 1.
 SO SEQUENCE 157 AA: 17370 MW: 6C6A837282EB826C CRC64;

Query Match 96.7%; Score 29; DB 5; Length 157;
 Best Local Similarity 80.0%; Pred. No. 1.9e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 DB 40 FPMWR 44

RESULT 4
 O9NMW3 PRELIMINARY: PRT; 158 AA.

AC 09NMW3; 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE MEMBRANE ANCHOR PROTEIN CD46 VARIANT (FRAGMENT)
 OS MCP.
 CN Homo sapiens (human).
 CC Eukaryota; Eukozoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBL_Taxid=9606;

RP SEQUENCE FROM N.A.
 RX MEDLINE=20227770; PubMed=10751138;
 RA Kusuhara K., Sasaki Y., Nakao F., Ihara K., Hattori H., Yamashita S.,
 RA Nihel K., Koide N., Alba H., Takeshita K., Hara T.;
 RT "Analysis of measles virus binding sites of the CD46 gene in patients
 with subacute sclerosing panencephalitis.";
 RL J. Infect. Dis. 181:1447-1449(2000).
 DR EMBL: AF209713; AAF73845.1; .

DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF00084; sushi_2.
 DR SMART: SM00032; CCP; 2.
 FT VARIANT 59 59 Q -> R.
 FT NON_TER 158 158
 SO SEQUENCE 158 AA: 18098 MW: 4E3F07EEC5C454F5 CRC64;

Query Match 96.7%; Score 29; DB 4; Length 158;

Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
11 11
DB 11 FPSWR 15

RESULT 5

ID 09NNM2 PRELIMINARY: PRT: 158 AA.

AC 09NNM2: 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE MEMBRANE COFACTOR PROTEIN CD46 VARIANT (FRAGMENT).
CN MCP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20227770; PubMed=10751138;
RA Kusuhara K., Sasaki Y., Nakao F., Ihara K., Hattori H., Yamashita S.,
RA Nishii K., Koide N., Alba H., Takeshita K., Hara T.;
RT "Analysis of measles virus binding sites of the CD46 gene in patients
RT with subacute sclerosing panencephalitis.";
RL J. Infect. Dis. 181:1447-1449(2000).
DR EMBL: AF209714; AAF73846.1; -;
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF00084; sushi. 2.
DR SMART: SM00032; CCP. 2.
FT NON_TER 158
FT SEQUENCE 158 AA; 18126 MW; E2CF38EC3A16A9DA CRC64;

Query Match 96.7%; Score 29; DB 4; Length 158;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
11 11
DB 11 FPSWR 15

RESULT 6
ID 037633 PRELIMINARY: PRT: 229 AA.

AC 037633: 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1) (FRAGMENT).
CN COIL.
OS Rhabdolepis boycei.
OC Mitochondrion.
OC Eukaryota; Metazoa; Artthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Tephritidae; Tephritidae; Rhabdolepis.
OX NCBI_TaxID=43419;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97159559; PubMed=9007018;
RA Smith J.J., Bush G.L.;
RT "Phylogeny of the genus Rhabdolepis (Diptera: Tephritidae) inferred
RT from DNA sequences of mitochondrial cytochrome oxidase II.";
RL Mol. Phylogenet. Evol. 7:33-43(1997).
DE FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
DE CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 2
DE 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 1
DE TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
DE A CENTER TO THE BINUCLEAR CENTER OF THE CATALYTIC SUBUNIT 1 (BY

SIMILARITY).
-1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O + 4
CC FERRICYTOCHROME C.
CC -1- COFACTOR: COPPER A AND HEME GROUP.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE (BY SIMILARITY).

CC -1- SIMILARITY: TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
DR EMBL: U53254; AB50355.1; -;
DR HSSP: P08306; IAR1.
DR InterPro: IPR001505; COX2.
DR InterPro: IPR002429; Cyt_c-ox_2.
DR Pfam: PF00116; COX2. 1.
DR PRINTS: PR01166; CYCOXIDASEII.
DR ProDom: PD000131; COX2. 1.
DR PROSITE: PS00078; COX2. 1.
KW Copper; Inner membrane; Mitochondrion; Oxidoreductase;
KW Respiratory chain; Transmembrane.
FT METAL 159 159 COPPER A (PROBABLE).
FT METAL 163 163 COPPER A (PROBABLE).
FT METAL 167 167 COPPER A (PROBABLE).
FT METAL 207 207 COPPER A (PROBABLE).
FT NON_TER 229
FT SEQUENCE 229 AA; 26389 MW; 912D1B2736330698 CRC64;

Query Match 96.7%; Score 29; DB 8; Length 229;
Best Local Similarity 80.0%; Pred. No. 2.7e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
11 11
DB 78 FPSWR 82

RESULT 7
ID 09PDE1 PRELIMINARY: PRT: 245 AA.

AC 09PDE1: 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE HYPOTHETICAL PROTEIN XF1438.
CN XF1438.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reimach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Bionnes M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carer H.,
RA Coutinho N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohne M., Furlan L.R.,
RA Garner J.M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Klajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite J.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madela A.M.B.N., Madela H.M.F., Martino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Neto L.E.S.,
RA Nhalai A., Nobrega F.C., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmenti D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A., Jr., Pasquero J.B.,
RA Queiroz R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA da Rosa V.E., Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,

RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*."
 RL Nature 406:151-159(2000).
 DR EMBL: AE003974; AAF84247.1; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 245 AA: 27193 MW: 16C8ADB9FC490455 CRC64:

Query Match: 96.7%; Score 29; DB 2; Length 245;
 Best Local Similarity 80.0%; Pred. No. 3.1e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PPXMR 5
 DB 197 FPSMR 201

RESULT 8
 O9SUV9 PRELIMINARY: PRT: 273 AA.

AC O9SUV9;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE HYPOHETICAL. 31.2 KDA PROTEIN.
 GN F7K2.180 OR AT4G22600.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Medler H., Wambutt R., Bancroft I., Newes H.W.,
 RA Mayer K.F.X., Lemcke K., Schueller C.;
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Medler H., Wambutt R., Newes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL: AL033545; CAA22165.1; -;
 DR EMBL: AL161557; CAB79215.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 273 AA: 31248 MW: 7E644F68CE8EBB0 CRC64:

Query Match: 96.7%; Score 29; DB 10; Length 273;
 Best Local Similarity 80.0%; Pred. No. 3.1e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PPXMR 5
 DB 79 FPSMR 83

RESULT 9
 O9HTB3 PRELIMINARY: PRT: 292 AA.
 AC O9HTB3;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE HYPOTHETICAL PROTEIN PA5488.
 GN PA5488.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warriner P.,
 RA Hickey M.J., Brickman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Gader R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Labig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Saiter M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an
 opportunistic pathogen."
 RL Nature 406:959-964(2000).
 DR EMBL: AE004961; AAG0873.1; -;
 FR InterPro: IPR000504; RRM.
 DR PROSITE: PS00030; RRM_RNP.1; UNKNOWN.1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 292 AA: 32949 MW: 14610337E8CDDFD4 CRC64:

Query Match: 96.7%; Score 29; DB 2; Length 292;
 Best Local Similarity 80.0%; Pred. No. 3.3e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PPXMR 5
 DB 231 FPSMR 235

RESULT 10
 O9D6L7 PRELIMINARY: PRT: 321 AA.

AC O9D6L7;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE 2310076014RIK PROTEIN.
 GN 2310076014RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=C57BL/6J; TISSUE=TONGUE;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baladrelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamliya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzaletti J., Mombert P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Samoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL: AK010208; BAB26769.1; -;

DR MGD: MG1:1919189; 2310076014RLK.
 DR InterPro: IPR000847; HTH_LYSR.
 DR PROSITE: PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.
 SO SEQUENCE 321 AA; 35038 MW; 6E27BEC746552DFB CRC64;

Query Match
 Best Local Similarity 96.7%; Score 29; DB 11; Length 321;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXWR 5
 DB 44 FPSWR 48

RESULT 11
 ID 09APJ5 PRELIMINARY; PRT: 332 AA.
 AC 09APJ5;
 DT 01-JUN-2001 (TReMBLrel. 17, Created)
 DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE EMDO-1,4-BETA-D-GLUCANASE PRECURSOR.
 GN CEL8Y.
 OS Erwinia chrysanthemi.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Pectobacterium.
 OX NCBI_TaxID=556;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PY35;
 RA Cho S.J., Park S.R., Yun H.D.;
 RT Cloning and sequencing of a cel8Y gene of Pectobacterium chrysanthemi
 RT PY35.
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF282321; AAC49556.1;
 KW Signal.
 FT SIGNAL 1 23 POTENTIAL.
 SO SEQUENCE 332 AA; 37627 MW; A1B5D7B0CB20EE6 CRC64;

Query Match
 Best Local Similarity 96.7%; Score 29; DB 2; Length 332;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXWR 5
 DB 178 FPWR 182

RESULT 12
 ID 09L3G9 PRELIMINARY; PRT: 333 AA.
 AC 09L3G9;
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4).
 GN CELA.
 OS Erwinia rhapontici.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Erwinia.
 OX NCBI_TaxID=55212;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCPPB2989;
 RA Saari-Lahti H.T.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCPPB2989;
 RA Rieki R.;
 RT Members of the amylovora group of Erwinia are cellulolytic and

KT possess genes homologous to the type II secretion pathway.
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ276358; CAB89803.1;
 DR InterPro: IPR002037; Glyco_hydro_8.
 DR Pfam: PF01370; Glyco_hydro_8; 1.
 DR PRINTS: PR00735; GLYDRLASE8.
 KW Signal; Hydrolase; GLYCOSIDASE.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 333 ENDOGLUCANASE.
 SO SEQUENCE 333 AA; 37783 MW; D75CFE212302673A CRC64;

Query Match
 Best Local Similarity 96.7%; Score 29; DB 2; Length 333;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXWR 5
 DB 178 FPWR 182

RESULT 13
 ID 09FTM4 PRELIMINARY; PRT: 334 AA.
 AC 09FTM4;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE P0005A05.25 PROTEIN.
 GN Oryza sativa (Rice).
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartioidae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponebare(GA3) genomic DNA, chromosome 1, PAC
 RT clone:P0005A05."
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AP002863; BAB16921.1;
 DR InterPro: IPR000520; Exonuclease.
 DR InterPro: IPR000822; Znf-C2H2.
 DR Pfam: PF00929; Exonuclease; 1.
 DR Pfam: PF00096; Zf-C2H2; 1.
 DR SMART: SM00479; EXO11; 1.
 DR SMART: SM00355; Znf-C2H2; 2.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 1.
 KW DNA-binding; Zinc-finger.
 SO SEQUENCE 334 AA; 37172 MW; A34531D62570A88 CRC64;

Query Match
 Best Local Similarity 96.7%; Score 29; DB 10; Length 334;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXWR 5
 DB 301 FPWR 305

RESULT 14
 ID 015429 PRELIMINARY; PRT: 349 AA.
 AC 015429;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE CD46.
 GN Homo sapiens (Human).

Search completed: February 27, 2002, 11:50:14
Job time: 993 sec

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Hara T., Suzuki Y., Nagasawa S., Seya T.;
RT "cDNA cloning and characterization of human sperm CD46."
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: D84105; BAA12224.1; -
DR HSSP: P10998; 1VD.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF00084; sushi; 4.
DR SMART: SM00032; CCP; 4.
SQ SEQUENCE 349 AA: 39325 MW: 8EFCEDA30D3C818E CRC64:

Query Match 96.7%; Score 29; DB 10; Length 349;
Best Local Similarity 80.0%; Pred. No. 4e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
DB 11 FPSWR 15

RESULT 15
09SAAI PRELIMINARY: PRT: 417 AA.
AC 09SAAI;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE F25C20.9.
CN F25C20.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Ienz C., Liu S.,
RA Li J., Lee J.M., Kremenetskaia I., Luros J., Ngan L., Liu A.,
RA Gonzalez A., Altafi H., Araujo R., Chao Q., Conn L., Conway A.B.,
Dunn P., Hansen N., Huizar L., Kim C., Palm C., Rowley D., Shinn P.,
Walker M., Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
RT "Arabidopsis thaliana chromosome 1 BAC F25C20 sequence."
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Theologis A.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Theologis A.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC007296; AAD30247.1; -
SQ SEQUENCE 417 AA: 45985 MW: B86EBF06AC19A63D CRC64:

Query Match 96.7%; Score 29; DB 10; Length 417;
Best Local Similarity 80.0%; Pred. No. 4.7e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
DB 377 FPSWR 381

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using SW model

Run on: February 27, 2002, 11:36:07 ; Search time 132.19 Seconds
(without alignments)
0.851 Million cell updates/sec

Title: US-09-446-109a-15

Perfect score: 30

Sequence: 1 FPXMR 5

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database:

Issued_Patents_AA: *
1: /cgn2_6/ptodata/2/iaa/5a_COMB.pep: *
2: /cgn2_6/ptodata/2/iaa/5b_COMB.pep: *
3: /cgn2_6/ptodata/2/iaa/6a_COMB.pep: *
4: /cgn2_6/ptodata/2/iaa/6b_COMB.pep: *
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep: *
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	96.7	38	1	US-08-176-500-118
2	29	96.7	38	1	US-08-471-052A-118
3	29	96.7	38	1	US-08-189-331-118
4	29	96.7	38	2	US-08-471-939-118
5	29	96.7	38	2	US-08-471-800-118
6	29	96.7	38	2	US-08-471-068-118
7	29	96.7	241	4	US-08-823-120-1
8	29	96.7	324	2	US-08-528-057-46
9	29	96.7	370	2	US-08-528-057-42
10	29	96.7	372	1	US-08-202-056-5
11	29	96.7	372	1	US-08-076-093A-6
12	29	96.7	372	1	US-08-701-265-6
13	29	96.7	372	2	US-08-284-586-6
14	29	96.7	372	2	US-08-805-478-6
15	29	96.7	372	2	US-08-802-627A-6
16	29	96.7	372	2	US-08-801-238-6
17	29	96.7	372	2	US-08-801-228-6
18	29	96.7	372	3	US-09-104-296-6
19	29	96.7	372	3	US-08-982-493-8
20	29	96.7	372	2	US-08-528-057-44
21	29	96.7	372	2	US-08-528-057-42
22	29	96.7	384	4	US-08-139-195-2
23	29	96.7	384	4	5514787-2
24	28	93.3	57	1	US-08-370-225-29
25	28	93.3	57	1	US-08-370-225-30
26	28	93.3	57	1	US-08-461-859-29
27	28	93.3	57	1	US-08-461-859-30

28	28	93.3	57	5	PCT-US93-10069-29	Sequence 29, Appl
29	28	93.3	57	5	PCT-US93-10069-30	Sequence 30, Appl
30	28	93.3	298	2	US-08-061-636-3	Sequence 3, Appl
31	28	93.3	298	2	US-08-874-347-19	Sequence 19, Appl
32	28	93.3	298	3	US-09-093-522-19	Sequence 19, Appl
33	28	93.3	298	5	PCT-US94-05268-3	Sequence 3, Appl
34	28	93.3	373	2	US-08-846-762-13	Sequence 13, Appl
35	28	93.3	374	2	US-08-982-493-6	Sequence 6, Appl
36	28	93.3	437	2	US-09-031-059-1	Sequence 1, Appl
37	28	93.3	437	2	US-09-031-059-3	Sequence 3, Appl
38	28	93.3	760	1	US-08-195-152-2	Sequence 2, Appl
39	26	86.7	10	4	US-09-461-697-409	Sequence 409, App
40	26	86.7	12	4	US-09-461-697-407	Sequence 407, App
41	26	86.7	113	1	US-07-668-648-10	Sequence 10, Appl
42	26	86.7	113	2	US-08-429-998-10	Sequence 10, Appl
43	26	86.7	113	2	US-08-431-333-10	Sequence 10, Appl
44	26	86.7	113	5	PCT-US91-02321-10	Sequence 10, Appl
45	26	86.7	862	1	US-08-325-267A-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-176-500-118
Sequence 118, Application US/08176500
Patent No. 5498538
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
APPLICANT: Fowlkes, D. M.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 141
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/176,500
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/013,416
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-143
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-176-500-118
Query Match 96.7%; Score 29; DB 1; Length 38;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
DB 10 FPSMR 14

RESULT 2

US-08-471-052A-118
Sequence 118, Application US/08471052A
Patent No. 5625033
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 166
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,052A
FILING DATE: 06-JUNE-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-179
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-471-052A-118

Query Match 96.7%; Score 29; DB 1; Length 38;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FPXMR 5
DB 10 FPSMR 14

RESULT 3

US-08-189-331-118
Sequence 118, Application US/08189331
Patent No. 5747334
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents.
NUMBER OF SEQUENCES: 186
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/189,331
FILING DATE: Concurrently herewith
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-155
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-189-331-118

Query Match 96.7%; Score 29; DB 1; Length 38;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FPXMR 5
DB 10 FPSMR 14

RESULT 4

US-08-471-939-118
Sequence 118, Application US/08471939
Patent No. 5844075
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 141
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,939
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/013,416
FILING DATE: 01-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-143
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:

SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-471-939-118

Query Match 96.7%: Score 29; DB 2; Length 38;
Best Local Similarity 80.0%: Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
11 11
DB 10 FPSMR 14

RESULT 5
US-08-471-800-118
Sequence 118, Application US/08471800
Patent No. 5852167
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
APPLICANT: Fowlkes, D. M.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 141
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,800
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/013,416
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mistock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-143
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-471-800-118

Query Match 96.7%: Score 29; DB 2; Length 38;
Best Local Similarity 80.0%: Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
11 11
DB 10 FPSMR 14

RESULT 6
US-08-471-068-118

Sequence 118, Application US/08471068
Patent No. 5948635
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
APPLICANT: Fowlkes, D. M.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 186
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,068
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/189,331

ATTORNEY/AGENT INFORMATION:

NAME: Mistock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-155
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 118:

SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-471-068-118

Query Match 96.7%: Score 29; DB 2; Length 38;
Best Local Similarity 80.0%: Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
11 11
DB 10 FPSMR 14

RESULT 7
US-08-823-120-1

Sequence 1, Application US/08823120
Patent No. 6149919
GENERAL INFORMATION:
APPLICANT: Domenighini, Mario
APPLICANT: Rappuoli, Rino
TITLE OF INVENTION: Immunogenic Detoxified Mutants of
TITLE OF INVENTION: Cholera Toxin and of the Toxin B, Their Preparation and
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: California
COUNTRY: USA

Query Match 96.7%: Score 29; DB 2; Length 38;
Best Local Similarity 80.0%: Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
11 11
DB 10 FPSMR 14

2IP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/823,120
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/256,003
FILING DATE: 11-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0315,001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 241 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-823-120-1

Query Match 96.7%; Score 29; DB 4; Length 241;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
DB 169 FPMR 173

RESULT 8
US-08-528-057-46
Sequence 46, Application US/08528057
Patent No. 5846715
GENERAL INFORMATION:
APPLICANT: PURCELL, Damian F. J.
APPLICANT: RUSSELL, Sarah M.
APPLICANT: MCKENZIE, Ian F. C.
TITLE OF INVENTION: CD46 VARIANTS
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/528,057
FILING DATE: CONCURRENTLY HERewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/961,686
FILING DATE: 11-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU91/00199
FILING DATE: 10-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU P0133/90
FILING DATE: 11-MAY-1990

ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 17227/112 DACO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 324 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-528-057-46

Query Match 96.7%; Score 29; DB 2; Length 324;
Best Local Similarity 80.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
DB 11 FPMR 15

RESULT 9
US-08-528-057-42
Sequence 42, Application US/08528057
Patent No. 5846715
GENERAL INFORMATION:
APPLICANT: PURCELL, Damian F. J.
APPLICANT: RUSSELL, Sarah M.
APPLICANT: MCKENZIE, Ian F. C.
TITLE OF INVENTION: CD46 VARIANTS
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/528,057
FILING DATE: CONCURRENTLY HERewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/961,686
FILING DATE: 11-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU91/00199
FILING DATE: 10-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU P0133/90
FILING DATE: 11-MAY-1990
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 17227/112 DACO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 370 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-528-057-42

Query Match 96.7%; Score 29; DB 1; Length 370;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FPXMR 5
db 11 FPSMR 15

RESULT 10
US-08-202-056-5
Sequence 5, Application US/08202056
Patent No. 5440021
GENERAL INFORMATION:
APPLICANT: Chuntarapai, Anan
APPLICANT: Hebert, Caroline
APPLICANT: Kim, Kyung Jin
APPLICANT: Lee, James
TITLE OF INVENTION: Antibodies to Human IL-8 Type B Receptor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: palin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202,056
FILING DATE: 25-FEB-1994
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/677211
FILING DATE: 29-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B.
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: 706P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ. ID NO.: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-202-056-5

Query Match 96.7%; Score 29; DB 1; Length 372;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FPXMR 5
db 352 FPSMR 356

RESULT 11
US-08-076-093A-6
Sequence 6, Application US/08076093A
Patent No. 5343503
GENERAL INFORMATION:

APPLICANT: Chuntarapai, Anan
APPLICANT: Lee, James
APPLICANT: Hebert, Caroline
APPLICANT: Jin Kim, K.
TITLE OF INVENTION: Antibodies to Human PFAA Receptors
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

US-08-701-265-6
Sequence 6, Application US/08701265
Patent No. 5776457
GENERAL INFORMATION:
APPLICANT: Chuntarapai, Anan
APPLICANT: Lee, James
APPLICANT: Hebert, Caroline
APPLICANT: Jin Kim, K.
TITLE OF INVENTION: Antibodies to Human PFAA Receptors
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

US-08-076-093A-6
Sequence 6, Application US/08076093A
Patent No. 5343503
GENERAL INFORMATION:

Query Match 96.7%; Score 29; DB 1; Length 372;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FPXMR 5
db 352 FPSMR 356

RESULT 12
US-08-701-265-6
Sequence 6, Application US/08701265
Patent No. 5776457
GENERAL INFORMATION:
APPLICANT: Chuntarapai, Anan
APPLICANT: Lee, James
APPLICANT: Hebert, Caroline
APPLICANT: Jin Kim, K.
TITLE OF INVENTION: Antibodies to Human PFAA Receptors
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/701,265
FILING DATE: 22-AUG-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/076093
FILING DATE: 11-JUN-1993
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
PRIOR APPLICATION DATA: 07/677211
FILING DATE: 29-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-701-265-6

Query Match 96.7% Score 29; DB 1; Length 372;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
DB 352 FPSMR 356

RESULT 13
US-08-284-586-6
Sequence 6, Application US/08284586
Patent No. 5840856
GENERAL INFORMATION:
APPLICANT: Chuntarapai, Anan
APPLICANT: Lee, James
APPLICANT: Hebert, Caroline
APPLICANT: Jin Kim, K.
TITLE OF INVENTION: Antibodies to Human PFAA Receptors
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/284,586
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/076,093A
FILING DATE: 11-JUN-1993
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/677211
FILING DATE: 29-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-284-586-6

Query Match 96.7% Score 29; DB 2; Length 372;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
DB 352 FPSMR 356

RESULT 14
US-08-805-478-6
Sequence 6, Application US/08805478
Patent No. 5874543
GENERAL INFORMATION:
APPLICANT: Chuntarapai, Anan
APPLICANT: Lee, James
APPLICANT: Hebert, Caroline
APPLICANT: Jin Kim, K.
TITLE OF INVENTION: ANTIBODIES TO PFAA RECEPTOR
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/805,478
FILING DATE: 25-FEB-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/284586
FILING DATE: 10-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/076093
FILING DATE: 11-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: P07062P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:

LENGTH: 372 amino acids
 TYPE: Amino Acid
 TOPOLOGY: Linear
 US-08-805-478-6

Search completed: February 27, 2002, 11:36:07
 Job time: 146 sec

Query Match 96.7%; Score 29; DB 2; Length 372;
 Best Local Similarity 80.0%; Pred. No. 1.6e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 DB 352 FPSMR 356

RESULT 15

US-08-802-627A-6
 Sequence 6, Application US/08802627A
 Patent No. 5892017
 GENERAL INFORMATION:
 APPLICANT: Lee, James
 APPLICANT: Wood, William L.
 TITLE OF INVENTION: NUCLEIC ACID ENCODING PFAA RECEPTOR
 NUMBER OF SEQUENCES: 6
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genentech, Inc.
 STREET: 460 Point San Bruno Blvd
 CITY: South San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94080
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5 inch, 1.44 MB floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Winpatin (Genentech)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/802,627A
 FILING DATE: 19-Feb-1997
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/284586
 FILING DATE: 10-AUG-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/076093
 FILING DATE: 11-JUN-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/810782
 FILING DATE: 19-DEC-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Love, Richard B.
 REGISTRATION NUMBER: 34,659
 REFERENCE/DOCKET NUMBER: P0706P2P1D2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415/225-5530
 TELEFAX: 415/952-9881
 TELEX: 910/371-7168
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 372 amino acids
 TYPE: Amino Acid
 TOPOLOGY: Linear
 US-08-802-627A-6

Query Match 96.7%; Score 29; DB 2; Length 372;
 Best Local Similarity 80.0%; Pred. No. 1.6e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 DB 352 FPSMR 356

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:41:19 ; Search time 303.5 Seconds
(without alignments)
1.220 Million cell updates/sec

Title: US-09-446-109A-16

Perfect score: 30

Sequence: 1 FPXMR 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A_Genseq_1101:*

- 1: /SID58/gcgdata/geneseq/geneseqp/AA1980.DAT.*
- 2: /SID58/gcgdata/geneseq/geneseqp/AA1981.DAT.*
- 3: /SID58/gcgdata/geneseq/geneseqp/AA1982.DAT.*
- 4: /SID58/gcgdata/geneseq/geneseqp/AA1983.DAT.*
- 5: /SID58/gcgdata/geneseq/geneseqp/AA1984.DAT.*
- 6: /SID58/gcgdata/geneseq/geneseqp/AA1985.DAT.*
- 7: /SID58/gcgdata/geneseq/geneseqp/AA1986.DAT.*
- 8: /SID58/gcgdata/geneseq/geneseqp/AA1987.DAT.*
- 9: /SID58/gcgdata/geneseq/geneseqp/AA1988.DAT.*
- 10: /SID58/gcgdata/geneseq/geneseqp/AA1989.DAT.*
- 11: /SID58/gcgdata/geneseq/geneseqp/AA1990.DAT.*
- 12: /SID58/gcgdata/geneseq/geneseqp/AA1991.DAT.*
- 13: /SID58/gcgdata/geneseq/geneseqp/AA1992.DAT.*
- 14: /SID58/gcgdata/geneseq/geneseqp/AA1993.DAT.*
- 15: /SID58/gcgdata/geneseq/geneseqp/AA1994.DAT.*
- 16: /SID58/gcgdata/geneseq/geneseqp/AA1995.DAT.*
- 17: /SID58/gcgdata/geneseq/geneseqp/AA1996.DAT.*
- 18: /SID58/gcgdata/geneseq/geneseqp/AA1997.DAT.*
- 19: /SID58/gcgdata/geneseq/geneseqp/AA1998.DAT.*
- 20: /SID58/gcgdata/geneseq/geneseqp/AA1999.DAT.*
- 21: /SID58/gcgdata/geneseq/geneseqp/AA2000.DAT.*
- 22: /SID58/gcgdata/geneseq/geneseqp/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	29	96.7	10	AAV08336	O. hannah venom al
2	29	96.7	11	AAV08334	O. hannah venom al
3	29	96.7	38	AAV08491	Binding domain of
4	29	96.7	94	AAV06599	Rat sodium channel
5	29	96.7	94	AAV06599	Rat truncated sodi
6	29	96.7	135	AAV04918	Mycobacterium spec
7	29	96.7	177	AAV04918	Breast and ovarian
8	29	96.7	299	AAV52488	Helicobacter pylori
9	29	96.7	314	AAV93943	CD46 construct del
10	29	96.7	324	AAV15233	CD46 from clone pm
11	29	96.7	359	AAV03396	Cellulase. Cellul

12	29	96.7	372	13	AAV27793	New platelet facto
13	29	96.7	372	16	AAV92239	Chemokine superlam
14	29	96.7	372	16	AAV68813	Human lymphocyte p
15	29	96.7	372	21	AAV06644	Human Burkitt's ly
16	29	96.7	372	21	AAV90627	Human G protein-co
17	29	96.7	372	21	AAV90661	Human mutant G pro
18	29	96.7	373	12	AAV15230	CD46 from clone pm
19	29	96.7	376	17	AAV93939	CD46 wild-type. H
20	29	96.7	377	12	AAV15229	CD46 from clone pm
21	29	96.7	377	17	AAV93941	CD46 construct sub
22	29	96.7	377	17	AAV93941	CD46 construct sub
23	29	96.7	384	18	AAV27484	Human MCP. Homo s
24	29	96.7	384	16	AAV10924	Human membrane cof
25	29	96.7	384	16	AAV86316	Human CD46. Homo
26	29	96.7	418	21	AAV13589	Streptomyces globi
27	29	96.7	421	21	AAV58394	Lung cancer associ
28	29	96.7	421	22	AAV75528	Human colon cancer
29	29	96.7	497	20	AAV04972	Mycobacterium sp
30	29	96.7	520	22	AAV90739	C. glutamicum prote
31	29	96.7	589	20	AAV34524	Porphyromonas ging
32	29	96.7	590	21	AAV94928	Human secreted pro
33	29	96.7	593	20	AAV34523	Porphyromonas ging
34	29	96.7	606	20	AAV34393	Porphyromonas ging
35	29	96.7	611	22	AAV03762	CAB-2 chimeric pro
36	29	96.7	612	22	AAV94498	Human protein sequ
37	29	96.7	614	18	AAV20591	H. pylori inner me
38	29	96.7	621	22	AAV94049	Human protein sequ
39	29	96.7	634	22	AAV40125	Human polypeptide
40	29	96.7	634	22	AAV92742	Human protein sequ
41	29	96.7	634	22	AAV93478	Human protein sequ
42	29	96.7	634	22	AAV93120	Human protein sequ
43	29	96.7	1464	19	AAV79294	An antigen from de
44	28	93.3	42	22	AAV15108	Peptide #1542 enco
45	28	93.3	42	22	AAV27563	Peptide #1600 enco

ALIGNMENTS

RESULT 1

ID AAV08336 standard: Peptide: 10 AA.

AC AAV08336:

PT 19-JUL-1999 (first entry)

XX

XX O. hannah venom alpha-neurotoxin homologous peptide 5.

XX Snake: venom: King cobra; alpha-neurotoxin; toxin; analgesia; screening.

XX

XX Ophiophagus hannah.

XX

XX WO924055-A1.

XX

XX 20-MAY-1999.

XX

XX 03-NOV-1998; 98WO-SC00087.

XX

XX 06-NOV-1997; 97SG-0003972.

XX

XX (UYST-) UNIV SINGAPORE NAT.

XX

XX Gopaladrishtakone P, Gwee MCE, Kini RM, Pu XC, Wong PT;

XX WPI: 1999-327205/27.

XX

XX Snake neurotoxin derived peptides

XX

XX Claim 11: Page 43; 46pp; English.

XX

XX This invention describes novel peptide derivatives of alpha-neurotoxin,

XX derived from snake venom toxins of Ophiophagus hannah and which are

CC capable of inducing or facilitating analgesia. The peptides, optionally
 CC with a carrier and/or diluent, are used to induce analgesia in a subject.
 CC Antibodies derived from the products of the invention can be used to
 CC screen for subject peptides, e.g. for purification purposes.
 XX

SO Sequence 10 AA;

Query Match 96.7%; Score 29; DB 20; Length 10;
 Best Local Similarity 80.0%; Pred. No. 13;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 II II
 DB 3 fpcwr 7

RESULT 2

AAV08334 standard; Peptide; 11 AA.

AC AAV08334;

DT 19-JUL-1999 (first entry)

DE O. hannah venom alpha-neurotoxin homologous peptide 3.

XX Snake: venom: King cobra; alpha-neurotoxin; toxin; analgesia; screening.

XX Ophiophagus hannah.

PN W0924055-A1.

PD 20-MAY-1999.

PF 03-NOV-1998; 98MO-SG00087.

PR 06-NOV-1997; 97SG-0003972.

PA (UYSI-) UNIV SINGAPORE NAT.

PI Copaladishnakone P, Gwee MCE, Kini RM, Pu XC, Wong PT;

DR WPI; 1999-327205/27.

PT Snake neurotoxin derived peptides

PS Claim 9; Page 42; 46pp; English.

XX This invention describes novel peptide derivatives of alpha-neurotoxin,
 CC derived from snake venom toxins of Ophiophagus hannah and which are
 CC capable of inducing or facilitating analgesia. The peptides, optionally
 CC with a carrier and/or diluent, are used to induce analgesia in a subject.
 CC Antibodies derived from the products of the invention can be used to
 CC screen for subject peptides, e.g. for purification purposes.
 XX

SO Sequence 11 AA;

Query Match 96.7%; Score 29; DB 20; Length 11;
 Best Local Similarity 80.0%; Pred. No. 14;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 II II
 DB 3 fpcwr 7

RESULT 3

AAV08491 standard; Protein; 38 AA.

AC AAV08491;

XX 18-APR-1995 (first entry)
 DT Binding domain of a polystyrene-binding TSAR (SB.9-5).
 XX

DE TSAR: totally synthetic affinity reagent; synthetic; binding domain;
 KW effector domain; concatenated heterofunctional protein; linker;
 KW direct; rapid; detection; screening; treatment; monoclonal antibody;
 KW Sm antigen; mouse model; autoimmune disease; SLE; Mab;
 KW systemic lupus erythematosus; polystyrene; SB.
 XX

OS Synthetic.

PN W09418318-A.

PD 18-AUG-1994.

PF 01-FEB-1994; 94MO-US00977.

PR 01-FEB-1993; 93US-0013416.

PR 30-DEC-1993; 93US-0176500.

PR 31-JAN-1994; 94US-0189331.

XX (UYNC-) UNIV NORTH CAROLINA.

PI Fowlkes DM, Kay BK;

DR WPI; 1994-279739/34.

PT Identifying proteins or peptide(s) which bind a ligand - by
 PT screening a recombinant vector library expressing fusion proteins
 PT comprising a binding domain and an effector domain
 XX

PS Example 7.8; Page 120; 255pp; English.

XX AAR58491-507 are amino acid sequences of the binding domain of TSAR
 CC (Totally Synthetic Affinity Reagents) peptides. These particular
 CC peptides are encoded by the polystyrene (SB) binding phage. There
 CC is no apparent linear motif, but the peptides are rich in Trp, Tyr and
 CC Gly, poor in Arg, Val and Lys, and completely lack cysteine residues.
 CC The non-variable amino acids at the NH2 and COOH terminals are
 CC not shown. TSAR peptides are generated using generic oligonucleotides
 CC (see AAO70470-73 for examples). TSARs are concatenated heterofunctional
 CC proteins or peptides, comprising at least two functional regions - a
 CC binding domain with affinity for a ligand and a second effector peptide
 CC portion that is chemically or biologically active. They may further
 CC comprise a linker peptide between the 2 domains. The TSARs or compns.
 CC comprising a TSAR binding domain can be used in vivo to deliver a
 CC chemically or biologically active moiety, eg. metal ion, radioisotope,
 CC peptide, toxin or enzyme, to the specific target or on the cell. They
 CC can also replace the function of macromolecules eg. monoclonal or
 CC polyclonal antibodies and therefore circumvent the need for complex
 CC methods of hybridoma formation or in vivo antibody production. The TSARs
 CC are easily characterised and have designed activity allowing direct and
 CC rapid detection in a screening process.
 XX

SO Sequence 38 AA;

Query Match 96.7%; Score 29; DB 15; Length 38;
 Best Local Similarity 80.0%; Pred. No. 46;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 II II
 DB 10 fpcwr 14

RESULT 4

AAV06599 standard; Protein; 94 AA.

AC AAV06599;

XX 26-OCT-1999 (first entry)
 XX Rat sodium channel Nan splice variant C-terminal sequence.
 DE
 XX
 XX NaN; sodium channel; ion transport; rat; dorsal root ganglia;
 KM palm; paraesthesia; hyperexcitability; therapy.
 XX
 OS Rattus sp.
 XX
 PN MO9938889-A2.
 XX
 PD 05-AUG-1999:
 XX
 XX 29-JAN-1999: 99WO-0502008.
 PF
 XX 20-NOV-1998: 98US-0109402.
 PR
 XX 29-JAN-1998: 98US-0072990.
 XX
 PA (UYVA) UNIV YALE.
 XX
 PI Dib-Hajj S, Waxman S;
 XX
 DR WPI: 1999-479168/40.
 XX
 XX New isolated nucleic acids encoding sodium channels, used to develop
 PT products for treating acute or chronic pain or hyperexcitability
 PT phenomena
 XX
 PS Example 5: Page 35; 91pp: English.
 XX
 CC This is the C-terminal sequence of a splice variant of rat sodium
 CC channel Nan. This truncated variant comprises amino acids 1-1378
 CC of the Nan sequence given in AAY06596 and the present, 94-amino acid
 CC sequence at the C-terminal end. It thus lacks the 387 C-terminal
 CC amino acids of full-length Nan. The variant arises from the use
 CC of a cryptic donor splice site in exon 23 and a novel exon 23'
 CC which is located in Intron 23 of the rat Nan gene. Nan is a
 CC previously unidentified voltage-gated sodium channel protein that
 CC is preferentially expressed in dorsal root ganglia or trigeminal
 CC ganglia, and which produces a TTX-R sodium current. The invention
 CC provides methods for identifying agents that modulate Nan channel
 CC activity or expression and for using such agents to treat acute or
 CC chronic pain, paraesthesia and hyperexcitability phenomena.
 CC
 XX
 SO Sequence 94 AA:

Query Match 96.7%; Score 29; DB 20; Length 94;
 Best Local Similarity 80.0%; Pred. No. 1,1e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
 II II
 Db 52 fpawr 56

RESULT 5
 AAB20130
 ID AAB20130 standard; Protein: 94 AA.
 XX
 AC AAB20130:
 XX
 DT 30-APR-2001 (first entry)
 XX
 XX Rat truncated sodium channel C-terminal region.
 DE
 XX Sodium channel; Nan; rat; tetrodotoxin resistant; pain;
 KM paraesthesia; hyperexcitability; analgesic; vaccine; therapy;
 KM diagnosis.
 XX
 OS Rattus norvegicus.
 XX

PN MO200105831-A1.
 XX
 XX 25-JAN-2001.
 PD
 XX
 PF 14-JUL-2000: 2000WO-US19342.
 XX
 PR 16-JUL-1999: 99US-0354147.
 XX
 PA (UYVA) UNIV YALE.
 XX
 PI Dib-Hajj S, Waxman SG;
 XX
 DR WPI: 2001-103147/11.
 XX
 XX Nucleic acid molecules encoding human tetrodotoxin resistant sodium
 PT channels, useful for preventing, diagnosing and treating pain,
 PT paraesthesia and/or hyperexcitability phenomena -
 PT
 XX
 PS Example 5: Page 34; 162pp: English.
 XX
 CC The present sequence is that of the C-terminal portion of a
 CC truncated rat Nan variant. The variant lacks the 387 C-terminal
 CC amino acids of full-length rat Nan (see AAR20122), which are
 CC replaced by the present 94-amino acid polypeptide. The N-terminal
 CC 1378 amino acids of the truncated variant are identical to those
 CC of full-length Nan. The new sequence arises from the use of a
 CC cryptic donor splice site in exon 23 and a novel exon 23' which is
 CC located in Intron 23. Rat Nan is a novel tetrodotoxin resistant
 CC sodium channel Nan belonging to the a-subunit voltage-gated sodium
 CC channel protein family. These sodium channels underlie the
 CC generation and propagation of impulses in excitable cells such as
 CC neurons and muscle fibres. Preferential expression of Nan on
 CC sensory neurons innervating the body (dorsal root ganglia) and
 CC the face (trigeminal ganglia), but not on other neurons, makes
 CC it a very useful target for diagnostic and/or therapeutic uses in
 CC relation to acute and/or chronic pain pathologies. A claimed
 CC method of treating pain, paraesthesia and/or hyperexcitability,
 CC phenomena in a human or animal subject involves administering an
 CC agent that alters sodium current flow through Nan channels, or
 CC which modulates transcription or translation of Nan mRNA, in
 CC dorsal root ganglia or trigeminal neurons. Nan polypeptides can
 CC be obtained by recombinant expression, and used to treat disorders
 CC associated with decreased sodium channel expression, to screen for
 CC compounds that modulate sodium channel expression or activity,
 CC and to raise antibodies useful as diagnostic agents.
 CC
 XX
 SO Sequence 94 AA:

Query Match 96.7%; Score 29; DB 22; Length 94;
 Best Local Similarity 80.0%; Pred. No. 1,1e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
 II II
 Db 52 fpawr 56

RESULT 6
 AAY04918
 ID AAY04918 standard; Protein: 135 AA.
 XX
 AC AAY04918:
 XX
 DT 06-JUL-1999 (first entry)
 XX
 XX Mycobacterium species protein sequence 358.
 DE
 XX Secreted protein: Mycobacterium; primer: PCR; amplification; probe;
 KM hybridisation; detection; vaccine; immunisation; infection.
 KM
 XX
 OS Mycobacterium sp.
 XX

PN W09909186-A2.
 XX
 PD 25-FEB-1999.
 XX
 PF 14-AUG-1998: 98MO-FR01813.
 XX
 PR 11-SEP-1997: 97FR-0011325.
 PR 14-AUG-1997: 97FR-0010404.
 XX
 PA (INSP) INST PASTEUR.
 PI Gicquel B., Lim EM, Pellicic V, Portnoi D, Coquet de la Salmoniere Y;
 PI Guigueno A.;
 DR WPI: 1999-181045/15.
 DR N-PSDB: AAX34171.
 XX
 PT Mycobacterial DNA vectors containing reporter constructs - for
 PT Identifying coding or promoter sequences involved in
 PT Infection-associated protein expression
 PS Claim 32: Fig 35B: 309pp: French.
 XX
 CC Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted
 CC proteins from various Mycobacterium species microorganisms. The
 CC encoding nucleotide sequences can be used as primers and probes for
 CC methods for detecting and identifying mycobacteria, especially belonging
 CC to the M. tuberculosis complex. The encoded proteins can be used in
 CC vaccines for immunisation against a bacterial or viral infection.
 XX
 SO Sequence 135 AA:

Query Match 96.7%; Score 29; DB 20; Length 135;
 Best Local Similarity 80.0%; Pred. No. 1.5e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FPXMR 5
 11 11
 Db 66 fpawr 70

RESULT 7

ID AAB58892 standard: Protein: 177 AA.

AC AAB58892:

DT 27-MAR-2001 (first entry)

DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 600.

XX Human: breast cancer; ovarian cancer; cytostatic; immunosuppressive;
 KM neoplastic; neuroprotective; antiviral; antiallergic; hepatotropic;
 KM antidiabetic; antiinflammatory; antitumor; vulnecary; anticonvulsant;
 KM antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
 KM Addison's disease; allergy; autoimmune haemolytic anaemia;
 KM autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
 KM multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
 KM cardiovascular disorder; wound healing; neurological disease.
 XX

OS Homo sapiens.

PN W0200055173-A1.

PD 21-SEP-2000.

PF 08-MAR-2000: 2000MO-US05881.

PR 12-MAR-1999: 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;
 XX WPI: 2000-611515/58.
 DR N-PSDB: AAF21795.
 XX
 PT New human breast and ovarian cancer associated gene sequences and the
 PT polypeptides encoded by these genes, useful in the prevention,
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular
 PT disorders and neurological diseases -
 XX
 PS Claim 11: Page 1037-1038; 1299pp; English.
 XX
 CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
 CC associated with breast and ovarian cancer. Included in the invention are
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
 CC isolation and characterisation of the DNA and protein sequences of the
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist
 CC or antagonist sequences exhibit cytostatic; immunosuppressive;
 CC neoplastic; neuroprotective; antiviral; antiallergic; hepatotropic;
 CC antidiabetic; antiinflammatory; antitumor; vulnecary; anticonvulsant;
 CC antibacterial; antifungal; antiparasitic and cardiant activity. The
 CC polynucleotide and protein sequences are used in the diagnosis of cancer,
 CC particularly breast and ovarian cancer. The nucleic acid sequences,
 CC proteins, agonists and antagonists may also be used in the diagnosis,
 CC prevention and treatment of immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; cardiovascular disorders such as
 CC myocardial ischaemias; wound healing; neurological diseases such as
 CC cerebral anoxia and epilepsy; and infectious diseases.
 XX
 SO Sequence 177 AA:

Query Match 96.7%; Score 29; DB 21; Length 177;
 Best Local Similarity 80.0%; Pred. No. 2e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FPXMR 5
 11 11
 Db 22 fpawr 26

RESULT 8

ID AAB52488 standard: Protein: 299 AA.

AC AAB52488:

DT 23-FEB-2001 (first entry)

DE Helicobacter pylori bait polypeptide #6.

XX Helicobacter pylori; two-hybrid system; protein-protein interaction;
 KM bait polypeptide; gastric ulcer; antibacterial.
 XX

CS Helicobacter pylori.

PN W0200066722-A1.

PD 09-NOV-2000.

PF 14-APR-2000: 2000MO-IB00603.

PR 30-APR-1999: 99EP-0401066.

PA (HYBR-) HYBRIGENICS SA.

PI Legrain P, Selig L, Rain J;

DR WPI: 2000-687335/67.

DR N-PSDB: AAC97234.

XX A two-hybrid system for identifying compounds useful in the treatment
 PT of e.g. gastric ulcers comprises producing a collection of recombinant
 PT cell clones -

XX Example 5: Page 88-89; 267pp: English.

XX The present sequence is a bait polypeptide used in a Helicobacter
 CC pylori two-hybrid screen to identify protein-protein interactions.
 CC The method is used to identify a recombinant cell clone expressing a
 CC prey polypeptide which is capable of interacting with the bait
 CC polypeptide. The two hybrid system is useful for screening compounds
 CC for antibacterial activity. It may be used in the treatment of gastric
 CC ulcers. The polynucleotides are useful as amplification primers or
 CC specific detection probes. The polypeptides, vectors or host cells can
 CC be used as immunogens to produce mono- or polyclonal antibodies. The
 CC polynucleotides, polypeptides, antibodies, vectors, host cells or
 CC modulating agents can be used to produce a pharmaceutical composition.

SQ Sequence 299 AA:

Query Match 96.7%; Score 29; DB 21; Length 299;
 Best Local Similarity 80.0%; Pred. No. 3.3e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 ||||
 DB 89 fpswr 93

RESULT 9

AAR93943 ID AAR93943 standard; Protein; 314 AA.

XX .AAR93943;

XX 21-MAY-1996 (first entry)

XX CD46 construct delSCR2/subSCR3.

XX CD46; recombinant protein; short consensus repeat; SCR;
 KW regulator of complement activation; transgenic animal; pig;
 KM organ transplantation.

XX Synthetic.

XX WO9606937-A1.

XX 07-MAR-1996.

XX 30-AUG-1995; 95MO-AU00553.

XX 30-AUG-1994; 94AU-0007724.

XX (AUST-) AUSTIN RES INST.

XX Christiansen D, Loveland B, McKenzie JFC, Milland J;

XX WPI: 1996-160368/16.

XX N-PSDB; AAT17599.

XX Increasing prodn. of recombinant proteins, esp. CD46 - by reducing
 PT the amt. of A and/or T in an A and/or T rich region of encoding gene
 PT exon

XX Claim 12: Page 40-41; 60pp: English.

XX CD46 delSCR2/subSCR3 (AAR93943) is the product of a cDNA construct
 CC (AAT17599) obt'd. by splice overlap extension PCR of wild-type CD46
 CC cDNA (AAT17593). The A+T content of A+T-rich exon 5 of the gene,
 CC encoding the short consensus repeat (SCR) 3 of CD46, was lowered
 CC and SCR2 was deleted. This resulted in improved prodn. in

CC eukaryotic host cells, e.g. CHO-K1 and COS-7. CD46 delSCR2/subSCR3
 CC is used to prevent complement- or inflammation-mediated tissue
 CC damage, to improve immunity to tumours or viruses, to control
 CC fertilisation and to prevent spontaneous abortion. Expression in
 CC transgenic animals, esp. pigs, provides organs suitable for
 CC transplantation.

SQ Sequence 314 AA:

Query Match 96.7%; Score 29; DB 17; Length 314;
 Best Local Similarity 80.0%; Pred. No. 3.4e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 ||||
 DB 11 fpswr 15

RESULT 10

AAR15233 ID AAR15233 standard; Protein; 324 AA.

XX .AAR15233;

XX 17-MAR-1992 (first entry)

XX CD46 from clone pm5.8.

XX SCR; short consensus repeats; transmembrane; cytoplasmic;
 KW membrane cofactor protein; MCP.

XX Homo sapiens.

XX OS

XX FH

XX Peptide

XX Protein

XX Modified-site

XX Modified-site

XX Domain

XX WO9118097-A.

XX 28-NOV-1991.

XX 10-MAY-1991; 91MO-AU00199.

XX 11-MAY-1990; 90AU-0000133.

XX (UYME-) UNIV MELBOURNE.

XX Purcell DJ, Russell SM, McKenzie JFC;

XX WPI: 1991-369251/50.

XX N-PSDB; AAO14915.

XX New CD46 membrane co-factor protein variants - useful as probes
 PT to identify CD46 isoforms and for diagnosing spontaneous
 PT abortion, inhibiting immuno:response and treating leukaemia

XX Disclosure: Fig 1A and 3B; 77pp: English.

XX The sequence of the pm5.8 clone is identical to the pm5.1 clone in
 CC the portion encoding the NH2 leader and four SCR regions. The
 CC sequence after nucleotide 890 was different, and results from
 CC reading through of the cDNA into an intron sequence after the
 CC fourth SCR. The protein encoded by this new sequence encodes
 CC a 16 amino acid hydrophobic region, probably serving as a membrane

CC spanning sequence. The putative cytoplasmic tail would also
 CC contain a high proportion of charged amino acids.
 CC See also AAQ14915-25, AAQ15211-12 and AAR15457-59.
 CC XX
 SQ Sequence 324 AA;

Query Match 96.7%; Score 29; DB 12; Length 324;
 Best Local Similarity 80.0%; Pred. No. 3.5e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FPXMR 5
 11 11
 Db 11 fpswr 15

RESULT 11

AAR70396
 ID AAR70396 standard; Protein: 359 AA.

AC AAR70396;
 DT 30-APR-1991 (first entry)
 XX
 DE Cellulase.
 XX
 KM Cellulase; plasmid; transformation.
 XX
 OS Cellulomonas uda CH4 (FER.

XX JP62175178-A.

XX 31-JUL-1987.

XX 30-JAN-1986; 86JP-0016800.

XX 30-JAN-1986; 86JP-0016800.

XX (FNER-) SHIN ENERGY SOGO KAIHATSU KIKO.
 PA (NENR-) NENRYOYO ALCOHOL KAIHATSU CIJUTSU KENKYU KUMIAT.

XX WPI: 1987-253827/36.

DR N-PSDB: AAN70617.

XX Bio-engineering cellulase produ. - by preparing DNA chain,
 PT preparing plasmid contg. DNA chain in the gene expressible state,
 PT and transforming bacterial by plasmid

XX Disclosure; Fig 1(A-B); 12pp; Japanese.

XX Transformation E.coli JM103 (FERM P-8612) with the gene encoding
 CC this protein allows expression of the protein and easy and cheap
 CC prodn. of cellulase.

XX Sequence 359 AA;

Query Match 96.7%; Score 29; DB 8; Length 359;
 Best Local Similarity 80.0%; Pred. No. 3.9e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
 11 11
 Db 178 fpswr 182

RESULT 12

AAR27793
 ID AAR27793 standard; Protein: 372 AA.

XX AAR27793;

XX 12-MAR-1993 (first entry)

XX New platelet factor 4 receptor superfamily member PF4AR1L.
 DE
 XX IL-8R; G-protein coupled receptor family; rhodopsin superfamily;
 KW pro-inflammatory cytokine; 8rr.9.
 KW
 XX

OS Homo sapiens.

XX WO9217497-A.

XX 15-OCT-1992.

XX 23-MAR-1992; 92MO-US02317.

XX 29-MAR-1991; 91US-0677211.

XX 19-DEC-1991; 91US-0810782.

XX (GETH) GENENTECH INC.

XX Holmes WE, Lee J, Wood WT.

XX WPI: 1992-366191/44.

XX N-PSDB: AAQ37107.

XX Isolated human platelet factor 4 superfamily receptor

PT polypeptide and corresp. antibodies and DNA - useful as

PT diagnostic and screening agents, and for treating inflammation or

PT PF4AR-mediated disorders

XX Claim 7; Fig 5; 78pp; English.

XX The IL-8 receptor cDNA sequence was isolated (see AAQ29505) and a

CC 874bp sub-fragment of the coding sequence was used as a probe to

CC screen human cell line HL60 and human peripheral blood lymphocyte

CC cDNA libraries. Two new gene sequences were found that are clearly

CC related to the IL-8 receptor. One of these was contained in clone

CC 8rr.9 and is predicted to encode an amino acid sequence which is

CC 36% and 38% identical with the high and low affinity IL-8 receptor

CC sequences, respectively. See also AAQ37107.

XX Sequence 372 AA;

XX Query Match 96.7%; Score 29; DB 13; Length 372;
 Best Local Similarity 80.0%; Pred. No. 4e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
 11 11
 Db 352 fpswr 356

XX RESULT 13

XX AAR92239 standard; Protein: 372 AA.

XX AAR92239;

XX 26-MAR-1996 (first entry)

XX Chemokine superfamily receptor.

XX Interleukin; IL-8; inflammation; psoriasis; dermatitis;

XX Rheumatoid arthritis; inflammatory bowel disease;

XX Chronic lung inflammation; treatment; antibody;

XX affinity purification; detection.

XX Homo sapiens.

XX US5440021-A.

XX 08-AUG-1995.

PF 29-MAR-1991: 91US-0677211.
 XX
 PR 25-FEB-1994: 94US-0202056.
 PR 29-MAR-1991: 91US-0677211.
 XX
 PA (CHUN/) CHUNTHARAPAI A.
 PA (HEBE/) HEBERT C.
 PA (KIM/) KIM K J.
 PA (LEE/) LEE J.
 XX
 PI Chuntharapai A, Hebert C, Kim KJ, Lee J:
 DR WPI: 1995-283151/37.
 DR N-PSDB: AAQ99009.
 XX
 PT New antibodies against interleukin 8 type B receptor - used to treat
 PT or prevent inflammation, also for detecting receptor expression and
 PT purification.
 PS
 PS Example 2: Columns 49-52; 62pp; English.
 CC Antibodies directed against the interleukin-8 receptor B can be used
 CC to treat or prevent inflammation e.g. psoriasis, dermatitis,
 CC rheumatoid arthritis and particularly inflammatory bowel disease and
 CC chronic lung inflammation. When immobilised, these antibodies may
 CC be used to detect interleukin-8 receptor B expression in cells and
 CC tissues and for affinity purification of interleukin-8 receptor B
 CC from cells. This sequence is an additional chemokine superfamily
 CC receptor which was identified by probing lambda libraries of genomic
 CC DNA from a human monocyte-like cell line (L-60) and human peripheral
 CC blood lymphocytes using a large fragment of the interleukin-8 type
 CC A receptor DNA (See AAQ99006).
 CC
 SQ Sequence 372 AA:
 Query Match 96.7%; Score 29; DB 16; Length 372;
 Best Local Similarity 80.0%; Pred. No. 4e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;
 QY 1 FPXMR 5
 Db 352 fpswr 356
 RESULT 14
 AAR68813 standard; Protein: 372 AA.
 XX
 AC AAR68813;
 XX
 DT 18-JUL-1995 (first entry)
 XX
 DE Human lymphocyte PF4AR.
 XX
 KM Interleukin-8 receptor; IL-8 receptor; PF4AR;
 KM platelet factor superfamily receptor; lymphocyte; chemotactic;
 KM inflammation; inflammatory disease; arthritis; emphysema; cystic;
 KM fibrosis; colitis; bronchitis; meningitis; therapeutic.
 XX
 OS Homo sapiens.
 XX
 PN MO9428931-A.
 PD 22-DEC-1994.
 PF 07-JUN-1994: 94MO-US06380.
 PR 11-JUN-1993: 93US-0076093.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Chuntharapai A, Hebert C, Kim KJ, Lee J:

XX
 DR WPI: 1995-036114/05.
 DR N-PSDB: AAQ80522.
 XX
 PT Treatment of inflammatory disorders - by administering an
 PT antibody capable of binding a platelet factor 4 superfamily
 PT receptor polypeptide
 PS
 PS Disclosure: Page 56-58; 83pp; English.
 CC 2 PF4AR members were identified by probing lambda libraries from
 CC human monocyte-like cell line HL-60 and human peripheral blood
 CC lymphocytes using a large fragment of IL-8 receptor DNA (full
 CC sequence given in AAQ80520). The nucleotide sequences of the 2
 CC PF4ARs are given in AAQ80521 and AAQ80522, and their respective
 CC amino acid sequences in AAR68812 and AAR68813.
 CC
 SQ Sequence 372 AA:
 Query Match 96.7%; Score 29; DB 16; Length 372;
 Best Local Similarity 80.0%; Pred. No. 4e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;
 QY 1 FPXMR 5
 Db 352 fpswr 356
 RESULT 15
 AAY06644 standard; Protein: 372 AA.
 XX
 AC AAY06644;
 XX
 DT 26-OCT-1999 (first entry)
 XX
 DE Human Burkitt's lymphoma receptor 1 (BLR1).
 XX
 KM Burkitt's lymphoma receptor 1; BLR1; human;
 KM B lymphocyte chemottractant; BLC; chemokine; ligand;
 KM drug screening; leukaemia; autoimmune disease; therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO9928468-A1.
 PD 10-JUN-1999.
 PF 02-DEC-1998: 98WO-US25561.
 PR 02-DEC-1997: 97US-0982493.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Cyster JG, Chun MD, Williams LT;
 DR WPI: 1995-03764/41.
 DR N-PSDB: AAX87710.
 XX
 PT Modulating interaction of a Burkitt's lymphoma Receptor 1
 PT polypeptide and ligand, useful in drug screens
 PS
 PS Claim 1: Page 38; 42pp; English.
 CC This sequence represents human Burkitt's lymphoma receptor 1
 CC (BLR1). The invention relates to methods for modulating the
 CC interaction of BLR1 with its ligand, B lymphocyte chemottractant
 CC (BLC, see AAY06642). The methods comprise combining BLR1 and BLC
 CC polypeptides with a candidate modulator agent under conditions
 CC whereby, but for the presence of the agent, the polypeptides engage
 CC in a first interaction, and determining a second interaction of the
 CC polypeptides in the presence of the agent, wherein a difference

CC between the first and second interactions indicates that the agent
CC modulates the interaction of the polypeptides. The modulator is
CC preferably an antagonist, especially dominant negative, form of
CC BLC. BLC and BLC agonists and antagonists may be useful for
CC treating viral (e.g. HIV) infections, lymphoma, B lineage
CC leukaemia, and autoimmune diseases such as rheumatoid arthritis,
CC thyroiditis and diabetes.

XX
SQ Sequence 372 AA:

Query Match: 96.7%; Score 29; DB 20; Length 372;

Best Local Similarity 80.0%; Pred. No. 4e+02; Mismatches 0; Caps 0;

Matches 4; Conservative 0; Indels 0; Caps 0;

OY 1 FPXMR 5

11 11

Db 352 fpwr 356

Search completed: February 27, 2002, 11:41:19
Job time: 458 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Comphen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:45:23 : Search time 145.23 seconds
(without alignments)
2.623 Million cell updates/sec

Title: US-09-446-109a-16

Perfect score: 30
Sequence: 1 FPXWR 5

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:
2: p1r2:
3: p1r3:
4: p1r4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	96.7	73	1 N2OH1	long neurotoxin 1
2	29	96.7	113	2 T07855	translation elonga
3	29	96.7	157	2 T24380	hypothetical prote
4	29	96.7	245	2 D82880	conserved hypothet
5	29	96.7	256	2 D85510	probable EC 3.5. a
6	29	96.7	259	2 A29831	heat-labile entero
7	29	96.7	273	2 T05454	hypothetical prote
8	29	96.7	292	2 C82960	hypothetical prote
9	29	96.7	327	2 S56162	MDCK15 protein - h
10	29	96.7	332	2 J07585	minor endoglucanas
11	29	96.7	349	2 C02913	sperm CD46 - human
12	29	96.7	359	2 I40696	endoglucanase - Ce
13	29	96.7	369	2 I57998	membrane cofactor
14	29	96.7	372	2 S26667	C protein-coupled
15	29	96.7	377	2 I34479	membrane cofactor
16	29	96.7	384	2 S01896	membrane cofactor
17	29	96.7	417	2 D86251	protein F25C20.9 (
18	29	96.7	479	2 T35441	aldehyde dehydroge
19	29	96.7	497	2 D83628	probable aldehyde
20	29	96.7	555	2 F69312	heterodisulfide re
21	29	96.7	625	2 H70310	hypothetical prote
22	29	96.7	642	2 D71909	ferrous iron trans
23	29	96.7	642	2 G64605	iron(II) transport
24	29	96.7	650	2 T22002	hypothetical prote
25	29	96.7	735	2 A83006	hypothetical prote
26	29	96.7	1451	2 I40325	demonstrated toxic
27	29	96.7	68	2 S75058	transposase ssr289
28	29	96.7	84	2 S76091	hypothetical prote
29	29	96.7	84	2 S76443	hypothetical prote

30	28	93.3	93	2 S75008	transposase ss1192
31	28	93.3	103	2 T47718	hypothetical prote
32	28	93.3	114	2 S77061	transposase sl1066
33	28	93.3	143	1 HHKW48	heat shock protein
34	28	93.3	143	1 HHKW41	heat shock protein
35	28	93.3	147	2 A71560	hypothetical prote
36	28	93.3	149	2 T35846	probable integral
37	28	93.3	157	2 F81710	conserved hypothet
38	28	93.3	169	2 G69300	conserved hypothet
39	28	93.3	254	2 S76459	transposase sl1186
40	28	93.3	254	2 E82791	conserved hypothet
41	28	93.3	258	2 S45682	acetone-cyanhydrin
42	28	93.3	259	2 S76643	transposase slr051
43	28	93.3	261	2 S77171	transposase sl1171
44	28	93.3	261	2 S77351	transposase sl1171
45	28	93.3	261	2 S75081	transposase slr026

ALIGNMENTS

RESULT 1
N2OH1
Long neurotoxin 1 - king cobra
N:Alternate names: neurotoxin A
C:Species: Ophiophagus hannah (king cobra)
C:Date: 22-Jun-1981 #sequence_revision 22-Jun-1981 #text_change 23-Aug-1996
C:Accession: A01658
R:Joubert, F.J.
Biochim. Biophys. Acta 317, 85-98, 1973
A:Title: Snake venom toxins. The amino acid sequences of two toxins from Ophiophagus
A:Reference number: A90588; MUID:73231298
A:Accession: A01658
A:Molecule type: protein
A:Residues: 1-73 <300>
C:Superfamily: snake toxin
C:Keywords: neurotoxin; venom
E:3-21,14-42,27-31,46-57,58-63/Disulfide bonds: #status predicted

Query Match 96.7% Score 29; DB 1; Length 73;
Best Local Similarity 80.0% Pred. No. 33;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 FPXWR 5
nb 66 FPXWR 70

RESULT 2
T07855
translation elongation factor eEF-1 alpha chain - rape (fragment)
N:Alternate names: translation elongation factor EF-1A
C:Species: Brassica napus (rape)
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 31-Mar-2000
C:Accession: T07855
R:Saenz-Vasquez, J.
submitted to the EMBL Data Library, February, 1995
A:Reference number: Z16177
A:Accession: T07855
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-113 <SAB>
X:Cross-references: EMBL:021744; NID:9914912; PIRN:AA86366.1; PID:9914913
A:Experimental source: cv. Samourai; 5 days old etiolated seedlings
C:Superfamily: translation elongation factor Tu; translation elongation factor Tu hom
C:Keywords: GTP binding; protein biosynthesis

Query Match 96.7% Score 29; DB 2; Length 113;
Best Local Similarity 80.0% Pred. No. 50;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 Db 103 FPMWR 107

RESULT 3

T24380
 hypothetical protein T03D8.2 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T24380
 R:Morimoto, B.
 submitted to the EMBL Data Library, March 1997
 A:Reference number: Z19884
 A:Accession: T24380
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-157 <MIL>
 A:Cross-references: EMBL:Z92838; PIDN:CAB07406.1; GSPDB:GN00023; CESP:T03D8.2
 A:Experimental source: clone T03D8
 C:Genetics:
 A:Gene: CESP:T03D8.2
 A:Map position: 5
 A:Introns: 49/3; 83/1; 125/3

Query Match 96.7%; Score 29; DB 2; Length 157;
 Best Local Similarity 80.0%; Pred. No. 68;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 Db 40 FPMWR 44

RESULT 4

D82680
 conserved hypothetical protein Xr1438 [imported] - *Xylella fastidiosa* (strain 9a5c)
 C:Species: *Xylella fastidiosa*
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: D82680
 R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
 A:Reference number: A82515; MUID:20365717
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: D82680
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-245 <SIM>
 A:Cross-references: GB:AE003974; GB:AE003849; NID:g9106454; PIDN:AAF84247.1; GSPDB:GN001
 A:Experimental source: strain 9a5c
 R:Simmons, A.J.G.; Relnack, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
 Brlones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carver, I
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincanli, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Frohm
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigi
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, R
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.T.;
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
 Rodrigues, V.; Rosa, A.J.; de M. de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
 M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: Xr1438

Query Match 96.7%; Score 29; DB 2; Length 245;
 Best Local Similarity 80.0%; Pred. No. 1e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 Db 197 FPMWR 201

RESULT 5

D85510
 probable EC 3.5. amidase-type enzyme yafV [imported] - *Escherichia coli* (strain O157:
 C:Species: *Escherichia coli*
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
 C:Accession: D85510
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
 Ilier, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Diallanta, E.; Polamoussis, K.; Apoda
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: D85510
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-256 <STO>
 A:Cross-references: GB:AE005174; NID:g12512987; PIDN:AAG54544.1; GSPDB:GN00145; UNCP:
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: yafV

Query Match 96.7%; Score 29; DB 2; Length 256;
 Best Local Similarity 80.0%; Pred. No. 1e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 Db 244 FPMWR 248

RESULT 6

A29831
 heat-labile enterotoxin Iia chain A precursor - *Escherichia coli*
 N:Alternate names: LT-Iia
 C:Species: *Escherichia coli*
 C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
 C:Accession: A29831
 R:Pickert, C.L.; Weinstein, D.L.; Holmes, R.K.
 J. Bacteriol. 169, 5180-5187, 1987
 A:Title: Genetics of type Iia heat-labile enterotoxin of *Escherichia coli*: operon *tus*
 A:Reference number: A91849; MUID:88032841
 A:Accession: A29831
 A:Molecule type: DNA
 A:Residues: 1-259 <PIC>
 A:Cross-references: GB:M17894; NID:g146671; PIDN:AAA24093.1; PID:g146672
 A:Note: the authors translated the codon TAT for residue 225 as Thr
 C:Superfamily: heat-labile enterotoxin chain A
 E:1-16/Domain: signal sequence #status predicted <STO>
 E:19-259/Product: heat-labile enterotoxin Iia chain A #status predicted <LTA>

Query Match 96.7%; Score 29; DB 2; Length 259;
 Best Local Similarity 80.0%; Pred. No. 1e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 Db 187 FPMWR 191

RESULT 7

T05454
 hypothetical protein F7K2.180 - *Arabidopsis thaliana*
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Sep-1999
 C:Accession: T05454
 R:Bevan, M.; Wedler, H.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schlue

submitted to the Protein Sequence Database, November 1998
A:Reference number: Z15416

A:Accession: T05454

A:Molecule type: DNA

A:Residues: 1-273 <REV>

A:Cross-references: EMBL:AL033545

A:Experimental source: cultivar Columbia; BAC clone F7K2

C:Genetics:

A:Map position: 4

A:Note: F7K2.180

C:Superfamily: Arabidopsis hypothetical protein F7K2.180

Query Match 96.7%; Score 29; DB 2; Length 273;

Best Local Similarity 80.0%; Pred. No. 1.1e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5

DB 79 FPSMR 83

RESULT 8

G82960 Hypothetical protein PA5488 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: G82960

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

..; Lory, S.; Olson, M.V.

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: A82950; MUID:20437337

A:Accession: G82960

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-292 <STO>

A:Cross-references: GB:AE004961; GB:AE004091; NID:g9951814; PIDN:AA08873.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA5488

Query Match 96.7%; Score 29; DB 2; Length 292;

Best Local Similarity 80.0%; Pred. No. 1.2e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5

DB 231 FPSMR 235

RESULT 9

S56162 MDCR15 protein - human

C:Species: Homo sapiens (man)

C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999

C:Accession: S56162

R:Barclay, L.; Loetscher, M.; Tobler, A.; Baggiolini, M.; Moser, B.

Biochem. J. 309, 773-779, 1995

A:Title: Sequence variation of a novel heptahelical leucocyte receptor through alternat

A:Reference number: S56162; MUID:95366951

A:Accession: S56162

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-327 <BAR>

A:Cross-references: EMBL:X68829; NID:g840783; PIDN:CAA48723.1; PID:g840784

C:Superfamily: vertebrate rhodopsin

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5

DB 307 FPSMR 311

RESULT 10

JT0585 minor endoglucanase (EC 3.2.1.-) precursor - Erwinia chrysanthemi

C:Species: Erwinia chrysanthemi

C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 08-Oct-1999

C:Accession: JT0585; JT0586

R:Guiseppi, A.; Aymeric, J.L.; Caml, B.; Barras, F.; Creuzet, N.

Gene 106, 109-114, 1991

A:Title: Sequence analysis of the cellulase-encoding cell gene of Erwinia chrysanthem

A:Reference number: JT0585; MUID:92039050

A:Accession: JT0585

A:Molecule type: DNA

A:Residues: 1-332 <CU1>

A:Cross-references: GB:M74044; NID:g148391; PIDN:AA24818.1; PID:g148392

A:Experimental source: strain 3937

A:Accession: JT0586

A:Molecule type: protein

A:Residues: 24-33 <CU11>

C:Genetics:

A:Gene: cely

C:Keywords: glycosidase; hydrolase

Fj:23/Domain: signal sequence #status predicted <SIG>

Fj:24-332/Product: minor endoglucanase #status predicted <MIN>

OY 1 FPXMR 5

DB 178 FPSMR 182

RESULT 11

G02913 sperm CD46 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 09-Jun-2000

C:Accession: G02913

K:Hara, T.

submitted to GenBank, March 1996

A:Reference number: H01942

A:Accession: G02913

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-349 <HAR>

A:Cross-references: GB:D84105; NID:g1256700; PIDN:BA12224.1; PID:g1256701

C:Superfamily: herpesvirus complement control protein; complement factor H repeat hom

Fj:35-94/Domain: complement factor H repeat homology <FH01>

Fj:99-157/Domain: complement factor H repeat homology <FH02>

Fj:162-223/Domain: complement factor H repeat homology <FH03>

Fj:228-283/Domain: complement factor H repeat homology <FH04>

Query Match 96.7%; Score 29; DB 2; Length 349;

Best Local Similarity 80.0%; Pred. No. 1.4e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5

DB 11 FPSMR 15

RESULT 12

I40696

endoglucanase - Cellulomonas uda
 C:Species: Cellulomonas uda
 C:Date: 04-Sep-1997 #sequence_rev1sion 04-Sep-1997 #text_change 05-Nov-1999
 C:Accession: 140696
 R:Nakamura, K.; Misawa, N.; Kitamura, K.
 J. Biotechnol. 4: 247-254, 1986
 A:Title: Sequence of a cellulase gene of Cellulomonas uda CB4.
 A:Reference number: 140696
 A:Accession: 140696
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-359 <RES>
 A:Cross-references: GB:M36503; NID:g144430; PIDN:AAA23090.1; PID:g144431

Query Match 96.7% Score 29; DB 2; Length 359;
 Best Local Similarity 80.0% Pred. No. 1.5e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 11 11
 DB 178 FPMWR 182

RESULT 13
 157998
 membrane cofactor protein - human
 C:Species: Homo sapiens (man)
 C:Date: 02-Jul-1996 #sequence_rev1sion 02-Jul-1996 #text_change 19-May-2000
 C:Accession: 157998
 R:Cervoni, F.; Fentich, P.; Akhoundi, C.; Hsli, B.L.; Rossi, B.
 Mol. Reprod. Dev. 34, 107-113, 1993
 A:Title: Characterization of a cDNA clone coding for human testis membrane cofactor protein
 A:Reference number: 157998; MUID:93119658
 A:Accession: 157998
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-369 <RES>
 A:Cross-references: GB:S51940; NID:g262937; PIDN:AA24802.1; PID:g262938
 C:Superfamily: herpesvirus complement control protein; complement factor H repeat homolo
 F:35-94/Domain: complement factor H repeat homology <FH01>
 F:99-157/Domain: complement factor H repeat homology <FH02>
 F:162-223/Domain: complement factor H repeat homology <FH03>
 F:228-283/Domain: complement factor H repeat homology <FH04>

Query Match 96.7% Score 29; DB 2; Length 369;
 Best Local Similarity 80.0% Pred. No. 1.5e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 11 11
 DB 11 FPMWR 15

RESULT 14
 S26667
 G protein-coupled receptor BLRI - human
 C:Species: Homo sapiens (man)
 C:Date: 13-Jan-1995 #sequence_rev1sion 13-Jan-1995 #text_change 21-Jul-2000
 C:Accession: S26667
 R:Dobner, T.; Wolf, I.; Emlich, T.; Lipp, M.
 Eur. J. Immunol. 22, 2795-2799, 1992
 A:Title: Differential expression of a novel G protein-coupled receptor from
 A:Reference number: S26667; MUID:93049615
 A:Accession: S26667
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-372 <DOH>
 A:Cross-references: EMBL:X68149; NID:g29459; PIDN:CAA48252.1; PID:g29460
 C:Genetics:
 A:Gene: GDB:BLRI
 A:Cross-references: GDB:136235; OMIM:601613

A:Map position: 15q26.1-15q26.1
 C:Superfamily: vertebrate rhodopsin
 C:Keywords: G protein-coupled receptor; transmembrane protein

Query Match 96.7% Score 29; DB 2; Length 372;
 Best Local Similarity 80.0% Pred. No. 1.5e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 11 11
 DB 352 FPMWR 356

RESULT 15
 154479
 membrane cofactor protein precursor, splice form pm5.1 - human
 R:Alternate names: lymphocyte surface glycoprotein CD46
 C:Species: Homo sapiens (man)
 C:Date: 07-Jun-1996 #sequence_rev1sion 07-Jun-1996 #text_change 31-Mar-2000
 C:Accession: 154479
 R:Porcull, D.F.; Russell, S.M.; Deacon, N.J.; Brown, M.A.; Hooker, D.J.; McKenzie, I.
 Immunogenetics 33, 335-344, 1991
 A:Title: Alternatively spliced RNAs encode several isoforms of CD46 (MCP), a regulato
 A:Reference number: 154479; MUID:91267562
 A:Accession: 154479
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-377 <RES>
 A:Cross-references: GB:M58050; NID:g180136; PIDN:AAA62833.1; PID:g180137
 C:Genetics:
 A:Gene: GDB:MCP
 A:Cross-references: GDB:120169; OMIM:120920
 A:Map position: 1q32-1q32
 C:Function:
 A:Description: for the factor I-mediated cleavage of the complement convertases
 C:Superfamily: herpesvirus complement control protein; complement factor H repeat hom
 C:Keywords: alternative splicing; glycoprotein; transmembrane protein
 F:35-94/Domain: complement factor H repeat homology <FH01>
 F:99-157/Domain: complement factor H repeat homology <FH02>
 F:162-223/Domain: complement factor H repeat homology <FH03>
 F:228-283/Domain: complement factor H repeat homology <FH04>
 F:329-351/Domain: transmembrane #status predicted <TM>

Query Match 96.7% Score 29; DB 2; Length 377;
 Best Local Similarity 80.0% Pred. No. 1.6e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 11 11
 DB 11 FPMWR 15

Search completed: February 27, 2002, 11:45:23
 Job time: 702 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:42:48 ; Search time 78.39 Seconds
(without alignments)
2.339 Million cell updates/sec

Title: US-09-446-109a-16
Perfect score: 30
Sequence: 1 FPXMR 5

Scoring table: BLOSUM62
Gapop 10.0, Capext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SWISSPROT_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	96.7	73	1 NXLI.OPHHA	P01387 ophiophagus
2	29	96.7	165	1 DSBV.VIBAL	Q56578 vibrio algi
3	29	96.7	259	1 E2AA.ECOLI	P13810 escherichia
4	29	96.7	332	1 CUNY.ERMCH	P27032 erwilia chr
5	29	96.7	359	1 GUN.CELUD	P18336 cellulomona
6	29	96.7	372	1 CCR5.HUMAN	P32302 homo sapien
7	29	96.7	377	1 MCP.HUMAN	P15529 homo sapien
8	29	96.7	420	1 2229.HUMAN	Q9UJW7 homo sapien
9	28	93.3	143	1 HS16.CAEBL	P06581 caenorhabdi
10	28	93.3	143	1 HS17.CAEBL	P02513 caenorhabdi
11	28	93.3	257	1 HNL.MANES	P52705 manihot esc
12	28	93.3	271	1 PABC.VIBHA	Q56693 vibrio harv
13	28	93.3	271	1 PBNK.BACST	P58053 bacillus st
14	28	93.3	286	1 PAL.KLEPN	P37446 klebsiella
15	28	93.3	289	1 PAL.ECOLI	P00631 escherichia
16	28	93.3	289	1 PAL.PROVU	P37447 proteus vul
17	28	93.3	289	1 PAL.SAITY	P37447 salmonella
18	28	93.3	298	1 CC2B.YEAST	P00546 saccharomyc
19	28	93.3	332	1 CC2.CAEBL	P34356 caenorhabdi
20	28	93.3	374	1 CCR5.MOUSE	Q04683 mus musculu
21	28	93.3	374	1 CCR5.RAT	P34997 rattus norv
22	28	93.3	430	1 KDTA.CHLMU	Q9PK15 chlamydia m
23	28	93.3	431	1 ANM2.HUMAN	Q57440 chlamydia t
24	28	93.3	433	1 ANM2.HUMAN	P55345 homo sapien
25	28	93.3	437	1 KDTA.CHLMU	Q46222 chlamydia p
26	28	93.3	448	1 Y347.CHLPN	Q92836 chlamydia p
27	28	93.3	459	1 RBL2.THIDE	Q60028 thlobacilli
28	28	93.3	463	1 RBL3.HYDMR	Q59462 hydrogenovi
29	28	93.3	470	1 LIP2.MYOCO	Q64424 myocastor c
30	28	93.3	482	1 CATFA.ONCYE	Q27710 onchocerca
31	28	93.3	490	1 CE05.ECOLI	Q47500 escherichia
32	28	93.3	490	1 CE10.ECOLI	Q41125 escherichia
33	28	93.3	492	1 CAT2.CUCPE	P48351 cucurbita p

34	28	93.3	492	1 CAT3.CUCPE	P48352 cucurbita p
35	28	93.3	548	1 CEAK.ECOLI	Q47502 escherichia
36	28	93.3	766	1 EYA.DROME	Q05201 drosophila
37	28	93.3	902	1 P2DH.RAT	P28037 rattus norv
38	28	93.3	973	1 TRP5.HUMAN	Q9U162 homo sapien
39	28	93.3	974	1 TRP4.MOUSE	Q9QU65 mus musculu
40	28	93.3	974	1 TRP5.RAHT	Q62852 oryctolagus
41	28	93.3	975	1 TRP5.MOUSE	Q9QU29 mus musculu
42	28	93.3	977	1 TRP4.HUMAN	Q9UJW4 homo sapien
43	28	93.3	977	1 TRP4.RAT	Q35119 rattus norv
44	28	93.3	981	1 TRP4.BOVIN	P79100 bos taurus
45	28	93.3	1056	1 DPOL.ADE02	P03261 human adeno

ALIGNMENTS

RESULT	ID	Sequence	Standard	PRT	73 AA
1	NXLI.OPHHA				
AC	P01387				
DT	21-JUL-1986 (Rel. 01, Created)				
DT	21-JUL-1986 (Rel. 01, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	LONG NEUROTOXIN 1 (NEUROTOXIN A).				
OS	Ophiophagus hannah (King cobra) (Naja hannah).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;				
OC	Elapidae; Elapinae; Ophiophagus.				
OX	NCBI_TaxID=8665;				
RN	[1]				
RP	SEQUENCE.				
RC	TISSUE=Venom;				
PX	MEDLINE=73231298; PubMed=4198767;				
FA	Joubert F.J.:				
RT	"Snake venom toxins the amino acid sequences of two toxins from				
RT	Ophiophagus hannah (King cobra) venom."				
RL	Biochim. Biophys. Acta 317:85-98(1973).				
CC	-1- SUBCELLULAR LOCATION: SECRETED.				
CC	-1- TISSUE SPECIFICITY: EXPRESSED BY VENOM GLAND.				
CC	-1- MISCELLANEOUS: LD(50) IS 0.3 MG/KG BY SUBCUTANEOUS INJECTION.				
CC	-1- SIMILARITY: BELONGS TO THE SNAKE TOXIN FAMILY.				
NR	PIR: A01658; N2OH1.				
NR	HSSP: P01386; ITRB.				
DR	Interpro: IPR003571; Snake_toxin.				
DR	Pfam: PF000087; toxin.1; Snake_toxin.				
DR	ProDom: PD000206; Snake_toxin.1.				
NR	ProSite: PS00272; SNAKE_TOXIN.1.				
KW	Venom; Neurotoxin; Multigene family.				
FT	DISULFID 3 21				
FT	DISULFID 14 42				
FT	DISULFID 27 31				
FT	DISULFID 46 57				
FT	DISULFID 58 63				
SO	SEQUENCE 73 AA: 8106 MW: 1AC17E91E16C54F7 CRC64:				

Query Match Score 29; DB 1; Length 73;

Best Local Similarity 80.0%; Pred. No. 17;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5

DB 66 FPXMR 70

RESULT 2

ISBB_VIBAL STANDARD: PRT: 165 AA.

AC 056578;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

```

DE DISULFIDE BOND FORMATION PROTEIN B (DISULFIDE OXIDOREDUCTASE)
DE (FRAGMENT).
DSBB.
OS Vibrio alginolyticus.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrrio.
RN NCBI_TaxID=663;
RP SEQUENCE FROM N.A.
RC STRAIN=B138-2;
RA Nakamura T., Enomoto H., Unemoto T.;
RL Submitted (Jul-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR DISULFIDE BOND FORMATION IN SOME
CC PERIPLASMIC PROTEINS. ACTS BY OXIDIZING THE DSB PROTEIN (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE DSBB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D83728; BAA12087.1; -.
DR InterPro: IPR003752; DsbB.
DR Pfam: PF02600; DsbB. 1.
KW Oxidoreductase; Redox-active center; Electron transport; Chaperone;
KW Transmembrane; Inner membrane.
FT DOMAIN 3 16 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 17 33 POTENTIAL.
FT DOMAIN 34 51 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 52 67 POTENTIAL.
FT DOMAIN 68 74 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 75 92 POTENTIAL.
FT DOMAIN 93 147 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 148 165 POTENTIAL.
FT DISULFID 43 46 REDOX-ACTIVE (BY SIMILARITY).
FT DISULFID 107 133 REDOX-ACTIVE (BY SIMILARITY).
FT NON_TER 165 165
SQ SEQUENCE 165 AA; 18745 MW; D43F96DF5AA3272D CRC64;

Query Match 96.7%; Score 29; DB 1; Length 165;
Best Local Similarity 80.0%; Pred. No. 37;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FPXWR 5
Db 113 FPSWR 117

RESULT 3
E2AA_ECOLI STANDARD; PRT; 259 AA.
AC P13810:
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HEAT-LABILE ENTEROTOXIN IIA, A CHAIN PRECURSOR (LT-IIA).
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN 111
RP SEQUENCE FROM N.A.
RC MEDLINE=88032841; PubMed=2822667;
RA Pickett C.L., Weinstein D.L., Holmes R.K.;
RT "Genetics of type IIA heat-labile enterotoxin of Escherichia coli:
RT operon fusions, nucleotide sequence, and hybridization studies.";
RL J. Bacteriol. 169:5180-5187(1987).

```

```

CC -1- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY
CC THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL. CYCLASE.
CC -1- SUBUNIT: HETEROHEXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M17894; AAA24093.1; -.
DR PIR: A29831; A29831.
DR HSSP: P43528; ITII.
DR InterPro: IPR001144; Enterotoxin_A.
DR Pfam: PF01375; Enterotoxin_A; 1.
DR PRINTS: PR00771; ENTEROTOXIN_A.
KW Enterotoxin; Signal.
FT SIGNAL 1 18
FT CHAIN 19 259 HEAT-LABILE ENTEROTOXIN IIA, A CHAIN.
FT DISULFID 203 215 BY SIMILARITY.
FT ACT_SITE 128 128 BY SIMILARITY.
SQ SEQUENCE 259 AA; 29242 MW; 996F311A332CABEA CRC64;

Query Match 96.7%; Score 29; DB 1; Length 259;
Best Local Similarity 80.0%; Pred. No. 56;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FPXWR 5
Db 187 FPAMR 191

RESULT 4
GUNT_ERMCH STANDARD; PRT; 332 AA.
AC P27032;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE MINOR ENDOGLUCANASE Y PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE
DE Y) (CELLULOSE Y) (ECY).
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=556;
RN 111
RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-33.
RC STRAIN=3937;
RA MEDLINE=92039050; PubMed=1937031;
RA Guisepi A., Aymeric J.-L., Caml B., Barras F., Creuzet N.;
RT "Sequence analysis of the cellulase-encoding cely gene of Erwinia
RT chrysanthemi: a possible case of interspecies gene transfer.";
RL Gene 106:109-114(1991).
CC -1- FUNCTION: REPRESENTS ONLY 34 OF THE GLOBAL CELLULOSE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: ENDOLYOLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY D (FAMILY 8 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M74044; AAA24819.1; -.
DR PIR: J70585; J70585.

```

DR InterPro: IPR002037; Glyco_hydro-8.
 DR Pfam: PF01270; Glyco_hydro-8; 1.
 DR PRINTS: PR00735; GLYDRLASE8.
 DR PROSITE: PS00812; GLYCOSYL_HYDROL_F8; 1.
 KW Cellulose degradation; Hydrolase; Glycosidase; signal.
 FT SIGNAL 1 23
 FT CHAIN 24 332 MINOR ENDOGLUCANASE Y.
 FT ACT_SITE 53 53 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 110 110 NUCLEOPHILE (POTENTIAL).
 SEQUENCE 332 AA: 37592 MW: 6EA760A2B8227079 CRC64;

Query Match: 96.7%; Score 29; DB 1; Length 332;
 Best Local Similarity 80.0%; Pred. No. 71;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1.FPKWR 5
 DB 178 FPAWR 182

RESULT 5
 CUN_CELUD STANDARD; PRT: 359 AA.
 AC P18336;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
 DE (CELLULOSE).
 OS Cellulomonas uda.
 OC Bacteria; Firmicutes; Actinobacteria; Actinomycetales;
 OC Actinomycetales; Micrococcales; Cellulomonadaceae; Cellulomonas.
 OX NCBI_TaxID=1714;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-37.
 RC STRAIN=CH4;
 RA Nakamura K., Misawa N., Kitamura K.;
 RT "Sequence of a cellulase gene of Cellulomonas uda CB4.";
 RL J. Biotechnol. 4:247-254(1986).

CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
 CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
 CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
 CC (2) EXOCELLULOBIODHOLASES THAT CUT THE DISSACCHARIDE CELLULOSE
 CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
 CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER
 CC SHORT CELLULO-OLIGOSACCHARIDES TO GLUCOSE.
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY D (FAMILY 8 OF GLYCOSYL
 CC HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL, outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL: M36503; AAA23090.1; -
 CC DR InterPro: IPR002037; Glyco_hydro-8.
 CC DR Pfam: PF01270; Glyco_hydro-8; 1.
 CC DR PROSITE: PS00812; GLYCOSYL_HYDROL_F8; 1.
 CC KW Cellulose degradation; Hydrolase; Glycosidase; signal.
 FT SIGNAL 1 23
 FT CHAIN 24 359 ENDOGLUCANASE.
 FT ACT_SITE 53 53 PROTON DONOR (BY SIMILARITY).
 FT ACT_SITE 110 110 NUCLEOPHILE (POTENTIAL).
 SEQUENCE 359 AA: 40690 MW: 0445D7571B683148 CRC64;

Query Match 96.7%; Score 29; DB 1; Length 359;

Best Local Similarity 80.0%; Pred. No. 76;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1.FPKWR 5
 DB 178 FPAWR 182

RESULT 6
 CCR5_HUMAN STANDARD; PRT: 372 AA.
 ID P32302; Q14811.
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE C-X-C CHEMOKINE RECEPTOR TYPE 5 (CXCR-5) (CXCR-5) (BURKITT'S LYMPHOMA
 DE RECEPTOR 1) (MONOCYTE-DERIVED RECEPTOR 15) (MDR15).
 GN BUR1 OR CXCR5.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG FORM).
 RC TISSUE=Lymphocytes;
 RX MEDLINE=93049615; PubMed=1425907;
 RA Dobner T., Wolf I., Emrich T., Hipp M.;
 RT "Differentiation-specific expression of a novel G protein-coupled
 RT receptor from Burkitt's lymphoma.";
 RL Eur. J. Immunol. 22:2795-2799(1992).
 RN [2]
 RP SEQUENCE FROM N.A. (SHORT FORM).
 RC TISSUE=Blood;
 RX MEDLINE=95366951; PubMed=7639692;
 RA Barrella L., Loetscher M., Tobler A., Baggiolini M., Moser B.;

RT "Sequence variation of a novel heptahelical leucocyte receptor
 RT through alternative transcript formation.";
 RL Biochem. J. 309:773-779(1995).
 RN [3]
 RP LIGAND BINDING.
 RX MEDLINE=98130629; PubMed=9463416;
 RA Legler D.F., Loetscher M., Stuber Roos R., Clark-Lewis I.,
 RA Baggiolini M., Moser B.;

RT "B cell-attracting chemokine 1, a human CXCR chemokine expressed in
 RT lymphoid tissues, selectively attracts B lymphocytes via
 RT BUR1/CXCR5.";
 RL J. Exp. Med. 187:655-660(1998).

CC -1- FUNCTION: CYTOKINE RECEPTOR THAT BINDS TO BLC. BUR1 EXERTS
 CC POSSIBLY A REGULATORY FUNCTION IN BURKITT LYMPHOMA (BL).
 CC LYMPHOMAGENESIS AND/OR B-CELL DIFFERENTIATION. IT IS A POTENTIAL
 CC CANDIDATE FOR CELL-CELL INTERACTION, AND ACTIVATION OF MATURE B-
 CC LYMPHOCYTES IN LYMPHATIC TISSUES.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM. ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -1- TISSUE SPECIFICITY: EXPRESSION IN MATURE B-CELLS AND BURKITT
 CC LYMPHOMA CELLS.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL, outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL: X68149; CAA48252.1; -
 CC DR EMBL: X68829; CAA48723.1; -
 CC DR PIR: S26667; S26667.
 CC DR GCRDB: GCR_0453; -
 CC DR GCRDB: GCR_2072; -
 CC DR GCRDB: GCR_2612; -

DR MIM: 601613; -
 DR InterPro: IPR000276; GPCR_Rhodopsn.
 DR Pfam: PF00001; 7tm_1; 1.
 DR PRINTS: PR00237; GPCRHHODPSN.
 DR PRINTS: PR00564; BURKILTSLYMR.
 DR PROSITE: PS00237; G_PROTEIN_RECEP_F1_1; 1.
 DR PROSITE: PS00262; G_PROTEIN_RECEP_F1_2; 1.
 DR G-protein coupled receptor; Transmembrane; Glycoprotein; B-cell;
 KW Alternative splicing.
 FT DOMAIN 1 55 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 56 76 1 (POTENTIAL).
 FT DOMAIN 77 88 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 89 109 2 (POTENTIAL).
 FT DOMAIN 110 124 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 125 145 3 (POTENTIAL).
 FT DOMAIN 146 167 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 168 188 4 (POTENTIAL).
 FT DOMAIN 189 219 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 220 240 5 (POTENTIAL).
 FT DOMAIN 241 259 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 260 280 6 (POTENTIAL).
 FT DOMAIN 281 304 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 305 325 7 (POTENTIAL).
 FT DOMAIN 326 372 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 28 28 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DISULFID 122 202 POTENTIAL.
 FT VARSLIC 1 45 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 344 344 C -> S (IN REF. 2).
 FT SEQUENCE 372 AA: 41955 MM: 6DF84C839492ACCF CRC64;

Query Match 96.7%; Score 29; DB 1; Length 372;
 Best Local Similarity 80.0%; Pred. No. 79;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 1 1
 1 1
 DB 352 FPSWR 356

RESULT 7
 MCP_HUMAN STANDARD: PRT: 377 AA.
 ID MCP_HUMAN P15529;
 AC P15529;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE MEMBRANE COFACTOR PROTEIN PRECURSOR (CD46 ANTIGEN) (TROPHOBLAST
 DE LEUCOCYTE COMMON ANTIGEN) (TLX).
 GN MCP.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxId=9606;
 RN 11
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 35-58.
 RX MEDLINE-88286080; PubMed-3260937;
 RX Lublin D.M., Liszewski M.K., Post T.W., Arce M.A., Le Beau M.M.,
 RA Rebertilich M.B., Lemons R.S., Seya T., Atkinson J.P.,
 RT Molecular cloning and chromosomal localization of human membrane
 RT cofactor protein (MCP). Evidence for inclusion in the multigene
 RT family of complement-regulatory proteins.*;
 RL J. Exp. Med. 168:181-194(1988).
 RN 12
 RP SEQUENCE FROM N.A.
 RC TISSUE-Testis;
 RA MEDLINE-93119658; PubMed-8418811;
 RA Garvon F., Fenichel P., Akhoundi C., Hsi B.L., Rossi B.,
 RT Characterization of a cDNA clone coding for human testis membrane
 RT cofactor protein (MCP, CD46).*;
 RL Mol. Reprod. Dev. 34:107-113(1993).
 RN 13

RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
 RX MEDLINE-91267562; PubMed-2050389;
 RA Purcell D.F., Russell S.M., Deacon N.J., Brown M.A., Hooker D.J.,
 RA McKenzie I.F.;
 RT "Alternatively spliced RNAs encode several isoforms of CD46 (MCP), a
 RT regulator of complement activation.*;
 RL Immunogenetics 33:335-344(1991).
 RN 14
 RP SEQUENCE OF 1-34 FROM N.A.
 RX MEDLINE-94014356; PubMed-7691939;
 RX Cui W., Hourcade D., Post T., Greenlund A.C., Atkinson J.P.,
 RA Kumar V.;
 RT Characterization of the promoter region of the membrane cofactor
 RT protein (CD46) gene of the human complement system and comparison to
 RT a membrane cofactor protein-like genetic element.*;
 RL J. Immunol. 151:4137-4146(1993).
 RN 15
 RP ALTERNATIVE SPLICING.
 RX MEDLINE-92289809; PubMed-1601037;
 RX Russell S.M., Sparrow R.L., McKenzie I.F.C., Purcell D.F.J.;
 RT "Tissue-specific and allelic expression of the complement regulator
 RT CD46 is controlled by alternative splicing.*;
 RL Eur. J. Immunol. 22:1513-1518(1992).
 CC -!- FUNCTION: MCP IS AN IMPORTANT MEMBRANE PROTEIN FOR PROTECTING HOST
 CC CELLS FROM DAMAGE BY COMPLEMENT. MCP POSSESSES COFACTOR ACTIVITY
 CC FOR INACTIVATION OF C3B AND C4B BY SERUM FACTOR I. THIS WOULD
 CC ALLOW MCP TO REGULATE THE COMPLEMENT SYSTEM AT THE LEVEL OF THE C3
 CC CONVERTASES. MAY BE INVOLVED IN THE FUSION OF THE SPERMATOZOA WITH
 CC THE OOCYTE DURING FERTILIZATION. MAY ACT BY PREVENTING COMPLEMENT
 CC ACTIVATION AT THE FETO-MATERNAL INTERFACE ON THE
 CC SYNCYTOTROPHOBLAST LAYER OF PLACENTA.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST 14 FORMS ARE PRODUCED BY
 CC ALTERNATIVE SPLICING. THE FORM SHOWN IS THE TRANSCRIPT A. THE
 CC CANNA ISOFORM (TRANSCRIPTS A AND B) IS PREFERENTIALLY EXPRESSED IN
 CC EBV-B CELLS AND LEUKEMIC CELLS. THE 66 KDA ALPHA ISOFORM AND THE
 CC 56 KDA BETA ISOFORM FOUND IN ALL TISSUES EXCEPT SPERM CORRESPOND
 CC TO THE TRANSCRIPTS C AND D, AND THE TRANSCRIPTS E AND F
 CC RESPECTIVELY. THE EXON 9 IS SPECIFICALLY DELETED IN SOME
 CC PLACENTA. THE EPSILON ISOFORM CORRESPONDS TO THE TRANSCRIPTS I
 CC AND J. SPERMATOZOA DELTA ISOFORM (TRANSCRIPT N) DELETE BOTH EXONS
 CC 12 AND 13. ALL TISSUES DIFFERENTIALLY SPLICE EXON 13.
 CC -!- TISSUE SPECIFICITY: EXPRESSED BY ALMOST ALL CELLS EXCEPT
 CC ERYTHROCYTES AND SOME BONE MARROW CELLS.
 CC -!- PTM: EXTENSIVELY O-GLYCOSYLATED IN THE SER/THR-RICH DOMAIN
 CC (PROBABLE).
 CC -!- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
 CC -!- DATABASE: NAME-PROW; NOTE-CD guide Cd46 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd46.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Y00651; CNA68675.1; -
 DR EMBL: S51940; AAB24802.1; -
 DR EMBL: M58050; AAA62833.1; -
 DR EMBL: A18585; CAA01400.1; -
 DR EMBL: S65879; AAD13968.1; -
 DR PIR: S01896; S01896.
 DR HSSP: P10998; LVVC.
 DR MIM: 120920; -
 DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF00084; sushi; 4.
 DR SMART: SM00032; CCP; 4.
 KW Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
 SUSHI: Alternative splicing.
 FT SIGNAL 1 34
 FT CHAIN 35 377 MEMBRANE COFACTOR PROTEIN.

```

FT DOMAIN 35 328 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 329 351 POTENTIAL.
FT DOMAIN 352 377 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 35 95 SUSHI 1.
FT DOMAIN 98 158 SUSHI 2.
FT DOMAIN 161 224 SUSHI 3.
FT DOMAIN 227 284 SUSHI 4.
FT DOMAIN 287 311 SER/THR-RICH.
FT DISULFID 35 80 BY SIMILARITY.
FT DISULFID 64 94 BY SIMILARITY.
FT DISULFID 99 141 BY SIMILARITY.
FT DISULFID 127 157 BY SIMILARITY.
FT DISULFID 162 210 BY SIMILARITY.
FT DISULFID 191 223 BY SIMILARITY.
FT DISULFID 228 270 BY SIMILARITY.
FT DISULFID 256 283 BY SIMILARITY.
FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 273 273 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPIC 286 300 MISSING (IN A SECOND FORM).
FT VARSPIC 340 352 MISSING (IN ISOFORM M).
FT VARSPIC 353 361 YLORRRKG -> DIEGGRKKGQWELNPLRLNQLQ
OSRRAE (IN ISOFORM M).
FT VARSPIC 362 377 TYLDETHREVKETSL -> KADGCAEYATYQTKSTTPAQ
RG (IN ISOFORM B/D/F/H/J/L).
FT VARSPIC 362 377 MISSING (IN ISOFORM M AND ISOFORM N).
FT VARSPIC 340 361 MISSING (IN ISOFORM N).
FT VARSPIC 339 339 I -> IGKQWELNPLRLNQLQOSRAE (IN
ISOFORM N).
SQ SEQUENCE 377 AA: 42247 MW: 2CA6F61752570B57 CRG64:

```

```

Query Match 96.7%: Score 29; DB 1: Length 377;
Best Local Similarity 80.0%: Pred. No. 80;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

```

```

QY 1 FPXWR 5
DB 11 FPSWR 15

```

```

RESULT 8
ID 2229_HUMAN STANDARD: PRT: 420 AA.
AC G9UJW7;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE ZINC FINGER PROTEIN 229 (FRAGMENT).
GN ZNF229.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Shannon M., Branscomb E., Hauser L., Gordon L., Ashworth L.,
RA Strubs L.;
RT "Differential expansion of homologous zinc-finger gene families in
human chromosome 19q13.2 and mouse chromosome 7."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Kodoylani V., Ge Y., Krummel G.K., Kvistad E., Grable L.,
RA Severin J., Gordon L., Shannon M., Brower A., Olsen A.S., Smith L.M.;
RT "Sequence analysis of a 1mb region in 19q13.2 containing a zinc finger
gene cluster."
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF192979; AAC07964.1; -
DR EMBL: AC084239; AAC23970.1; -
DR InterPro: IPR001909; KRAB.
DR InterPro: IPR000822; Znf-C2H2.
DR Pfam: PF01352; KRAB.1.
DR Pfam: PF00096; zf-C2H2; 2.
DR PRINTS: PR00048; ZINCFINGER.
DR SMART: SM00349; KRAB.1.
DR SMART: SM00355; ZNF_C2H2; 2.
DR PROSITE: PS50805; KRAB.1.
DR PROSITE: PS50028; ZINC_FINGER_C2H2_1; 2.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 3.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 34 108 KRAB.
FT DOMAIN 349 >420 ZINC_FINGERS.
FT ZN_FING 349 371 C2H2-TYPE.
FT ZN_FING 377 399 C2H2-TYPE.
FT ZN_FING 405 >420 C2H2-TYPE.
FT NON_TER 420 420
SQ SEQUENCE 420 AA: 48022 MW: FA4138BA44A2A14 CRG64:

```

```

Query Match 96.7%: Score 29; DB 1: Length 420;
Best Local Similarity 80.0%: Pred. No. 89;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

```

```

QY 1 FPXWR 5
DB 173 FPWR 177

```

```

RESULT 9
ID HS16_CAEEL STANDARD: PRT: 143 AA.
AC P06581; P02514;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE HEAT SHOCK PROTEIN HSP16-41.
GN HSP16-41.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=86304344; PubMed=3017958;
RA Jones D., Russnak R.H., Kay R.J., Candido E.P.M.;
RT "Structure, expression, and evolution of a heat shock gene locus in
Caenorhabditis elegans that is flanked by repetitive elements."
RL J. Biol. Chem. 261:12006-12015(1986).
RN [2]
RP SEQUENCE OF 47-143 FROM N.A.
RA MEDLINE=83220736; PubMed=6190129;
RA Russnak R.H., Jones D., Candido E.P.M.;
RT "Cloning and analysis of cDNA sequences coding for two 16 kilodalton
heat shock proteins (hsps) in Caenorhabditis elegans: homology with
the small hsps of Drosophila."
RL Nucleic Acids Res. 11:3187-3205(1983).
CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch).

DR EMBL: M14334; AAA28070.1; ALT_SEQ.
 DR EMBL: X01577; CAA25732.1; -
 DR PIR: A38884; HKW41.
 DR PIR: A25199; A25199.
 DR InterPro: IPR002068; Crystallin_HSP20.
 DR Pfam: PF00011; HSP20; 1.
 DR PROSITE: PS01031; HSP20; 1.
 DR Heat shock; Multigene family.
 KM SEQUENCE 143 AA; 16252 MW; CID0F59D26E36C74 CRC64;

Query Match 93.3%; Score 28; DB 1; Length 143;
 Best Local Similarity 80.0%; Pred. No. 50;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXWR 5
 11 11
 27 FPYWR 31

Db 27 FPYWR 31

RESULT 10
 HS17_CAEL STANDARD; PRT; 143 AA.
 ID HS17_CAEL
 AC P02513;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE HEAT SHOCK PROTEIN HSP16-48.
 GN HSP16-48.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Melazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Melazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 NCBI_TaxID=6239;
 RX SEQUENCE FROM N.A.
 MEDLINE=85295957; PubMed=403652;
 RA Russnak R.H., Candido E.P.M.;
 RT "Locus encoding a family of small heat shock genes in Caenorhabditis
 RT elegans: two genes duplicated to form a 3.8-kilobase inverted
 RT repeat".
 RL Mol. Cell. Biol. 5:1268-1278(1985).
 RN [2]
 RP SEQUENCE OF 9-143 FROM N.A.
 RX MEDLINE=8320736; PubMed=6190129;
 RA Russnak R.H., Jones D., Candido E.P.M.;
 RT "Cloning and analysis of cDNA sequences coding for two 16 kilodalton
 RT heat shock proteins (hsps) in Caenorhabditis elegans: homology with
 RT the small hsps of Drosophila".
 RL Nucleic Acids Res. 11:3187-3205(1983).
 CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch).

DR EMBL: K03273; AAA28069.1; -
 DR EMBL: X01576; CAA25731.1; -
 DR EMBL: K01863; AAA28064.1; -
 DR PIR: A02916; HKW48.
 DR PIR: A24289; A24289.
 DR InterPro: IPR002068; Crystallin_HSP20.

DR Pfam: PF00011; HSP20; 1.
 DR PROSITE: PS01031; HSP20; 1.
 KM Heat shock; Multigene family.
 KM SEQUENCE 143 AA; 16299 MW; 0D5596DFE5B3318 CRC64;

Query Match 93.3%; Score 28; DB 1; Length 143;
 Best Local Similarity 80.0%; Pred. No. 50;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXWR 5
 11 11
 27 FPYWR 31

Db 27 FPYWR 31

RESULT 11
 HNL_MANES STANDARD; PRT; 257 AA.
 ID HNL_MANES
 AC P52705;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE (S)-ACETONE-CYANOHYDRIN LYASE (EC 4.1.2.39) ((S)-HYDROXYNITRILE LYASE)
 DE ((S)-HYDROXYNITRILASE) (OXYNITRILASE).
 GN HNL.
 OS Manihot esculenta (Cassava) (Manioc).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; Rosidae;
 OC eudicots I; Malpighiales; Euphorbiaceae; Manihot.
 NCBI_TaxID=3983;
 RX SEQUENCE FROM N.A. AND SEQUENCE OF 1-36 AND 168-191.
 RP TISSUE-Cultured;
 RC MEDLINE=94263231; PubMed=8203915;
 RA Hughes J., Carvalho F.J.P.D.C., Hughes M.A.;
 RT "Purification, characterization, and cloning of alpha-hydroxynitrile
 RT lyase from cassava (Manihot esculenta Cratz)."
 RL Arch. Biochem. Biophys. 311:496-502(1994).
 CC -1- FUNCTION: INVOLVED IN CYANOGENESIS. THE RELEASE OF HCN FROM
 CC INJURED TISSUES. DECOMPOSES A VARIETIES OF (R) OR (S) CYANOHYDRINS
 CC INTO HCN AND THE CORRESPONDING ALDEHYDES AND KETONES. THE NATURAL
 CC SUBSTRATE OF THIS ENZYME IS (S)-ACETONE CYANOHYDRIN.
 CC -1- CATALYTIC ACTIVITY: (S)-2-HYDROXYISOBUTYRONITRILE = CYANIDE +
 CC ACETONE.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SIMILARITY: DISTANT RELATIONSHIP WITH TYPE-B
 CC CARBOXYLESTERASE/LIPASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch).

DR EMBL: Z29091; CAA82334.1; -
 DR HSSP: P52704; IVAS.
 DR InterPro: IPR000073; Abhydrolase.
 DR InterPro: IPR000379; Est_hlp_thioest_actsite.
 DR Pfam: PF00561; abhydrolase; 1.
 KM Lyase.
 FT INIT_MET 0
 FT ACT_SITE 79 79 BY SIMILARITY.
 FT ACT_SITE 207 207 BY SIMILARITY.
 FT ACT_SITE 235 235 BY SIMILARITY.
 SO SEQUENCE 257 AA; 29240 MW; 98B3E160ACB338C5 CRC64;

Query Match 93.3%; Score 28; DB 1; Length 257;
 Best Local Similarity 80.0%; Pred. No. 87;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
11 11
Db 124 FPDWR 128

RESULT 12

PABC_VIBHA STANDARD: PRT: 271 AA.
ID PABC_VIBHA
AC 056693;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE 4-AMINO-4-DEOXYCHORISMATE LYASE (EC 4.1.3.1) (ADC LYASE) (ADCL).
GN PABC.
OS Vibrio harveyi.
OC Bacteria: Proteobacteria: gamma subdivision: Vibrionaceae: Vibrio.
OX NCBI_TaxID=669;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN-B392;
RX MEDLINE=96134997; PubMed=8550484;
RA Shen Z., Biers D.M.;
RT *Isolation of Vibrio harveyi acyl carrier protein and the fabG, acpP, and fabE genes involved in fatty acid biosynthesis.*;
RL J. Bacteriol. 178:571-573(1996).
CC -1- FUNCTION: CONVERTS 4-AMINO-4-DEOXYCHORISMATE INTO 4-AMINO-BENZOATE (PABA) AND PYRUVATE.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -1- PATHWAY: SECOND STEP IN FOLATE BIOSYNTHESIS PATHWAY.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT AMINOTRANSFERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U39441; AAC43592.1; -
CC InterPro: IPR001544; AminoTran_4.
CC DR Pfam: PF01063; AminoTran_4; 1.
CC DR ProDom: PD001961; AminoTran_4; 1.
CC DR PROSITE: PS00770; AA_TRANSFER_CLASS_4; 1.
CC KM Lyase: Pyridoxal phosphate: Folate biosynthesis.
CC FT LYDING 140 140
CC FT PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC FT SEQUENCE 271 AA; 29958 MW; 3A5EB9F398D1CD CRC64;

Query Match 93.3%; Score 28; DB 1; Length 271;
Best Local Similarity 80.0%; Pred. No. 91;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
11 11
Db 57 FPDWR 61

RESULT 13

PPNK_RACST STANDARD: PRT: 271 AA.
ID PPNK_RACST
AC P58055;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE INORGANIC POLYPHOSPHATE/ATP-NAD KINASE (EC 2.7.1.23)
GN PPNK OR NADK.
OS Bacillus stearothermophilus.
OC Bacteria: Firmicutes: Bacillus/Clostridium group:
OC Bacillus/Staphylococcus group; Geobacillus.

OX NCBI_TaxID=1422;
RN 111
RP SEQUENCE FROM N.A.
RA Ohshima T., Sakuraba H.;
RT "Inorganic polyphosphate/ATP-NAD kinase of Bacillus stearothermophilus.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CATALYZES THE PHOSPHORYLATION OF NAD TO NADP. UTILIZES ATP AND OTHER NUCLEOSIDE TRIPHOSPHATES AS WELL AS INORGANIC POLYPHOSPHATE AS A SOURCE OF PHOSPHORUS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + NAD(+) -> ADP + NADP(+) (BY SIMILARITY).
CC -1- COFACTOR: REQUIRES DIVALENT METAL IONS FOR ACTIVITY (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOSOLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NAD KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AB055961; BAB32727.1; -
CC DR Transferase: Kinase: NAD: NADP.
CC KM EMBL: AB055961; BAB32727.1; -
CC FT SEQUENCE 271 AA; 30517 MW; 9E0FB427BC76662A CRC64;

Query Match 93.3%; Score 28; DB 1; Length 271;
Best Local Similarity 80.0%; Pred. No. 91;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
11 11
Db 258 FPDWR 262

RESULT 14

PAL_KLEPN STANDARD: PRT: 286 AA.
ID PAL_KLEPN
AC P37446;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PHOSPHOLIPASE A1 PRECURSOR (EC 3.1.1.32) (DIFFERENT-RESISTANT DE PHOSPHOLIPASE A) (DR-PHOSPHOLIPASE A) (PHOSPHATIDYLCHOLINE 1-ACYLHYDROLASE) (OUTER MEMBRANE PHOSPHOLIPASE A) (OM PLA).
GN PLDA.
OS Klebsiella pneumoniae.
OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
OC Klebsiella.
OX NCBI_TaxID=573;
RN 111
RP SEQUENCE FROM N.A.
RX MEDLINE=94131966; PubMed=8300539;
RA Brok R.G.P.M., Brinkman E., van Boxtel R., Bekkers A.C.A.P., Verheij H.M., Tommassen J.;
RT "Molecular characterization of enterobacterial plda genes encoding outer membrane phospholipase A.";
RL J. Bacteriol. 176:861-870(1994).
CC -1- FUNCTION: HYDROLYSIS OF PHOSPHATIDYLCHOLINE WITH PHOSPHOLIPASE A2 (EC 3.1.1.4) AND PHOSPHOLIPASE A1 (EC 3.1.1.32) ACTIVITIES.
CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O -> 1-ACYLGLYCEROPHOSPHOCHOLINE + A FATTY ACID ANION.
CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O -> 2-ACYLGLYCEROPHOSPHOCHOLINE + A FATTY ACID ANION.
CC -1- COFACTOR: REQUIRES CALCIUM IONS FOR ACTIVITY.
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE; ONE OF THE VERY FEW ENZYMES LOCATED THERE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: X76901; CA54223.1; -

DR PIR: B36971; B36971.

DR PIR: S40129; S40129.

DR InterPro: IPR003187; PLA1.

DR Pfam: PF02253; PLA1: 1

KM HydroLase: Lipid degradation: Outer membrane: Signal: Calcium.

FT SIGNAL 1 20 BY SIMILARITY.

FT CHAIN 21 286 PHOSPHOLIPASE A1.

FT ACT_SITE 161 161 BY SIMILARITY.

SO SEQUENCE 286 AA; 32544 MW; 3E39F863085108A3 CRC64;

Query Match 93.3%; Score 28; DB 1; Length 286;
 Best Local Similarity 80.0%; Pred. No. 96;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5

DB 92 FPLMR 96

RESULT 15

PAL_ECOLI STAMP: 289 AA.

AC P00631;

DT 21-JUL-1986 (Rel. 01, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE PHOSPHOLIPASE A1 PRECURSOR (EC 3.1.1.32) (DETERGENT-RESISTANT
 DE PHOSPHOLIPASE A) (DR-PHOSPHOLIPASE A) (PHOSPHATIDYLCHOLINE 1-
 DE ACYLHYDROLASE) (OUTER MEMBRANE PHOSPHOLIPASE A) (OM PLA).

GN PLPA OR B3821 OR Z5342 OR ECS4751.

OS Escherichia coli, and

OS Escherichia coli O157:H7.

OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:

OC Escherichia.

OX NCBI_TaxID=562, 83334;

RN 11

RP SEQUENCE FROM N.A.

RX MEDLINE=85157492; PubMed=6397464;

RA Homma H., Kobayashi T., Chiba N., Karasawa K., Mizushima H., Kudo T.,
 RA Inoue K., Ikeda H., Sekiguchi M., Nojima S.;

RT "The DNA sequence encoding plid gene, the structural gene for
 RT detergent-resistant phospholipase A of E. coli.";

RL J. Biochem. 96:1655-1664(1984).

RN 12

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;

RX MEDLINE=92358234; PubMed=1379743;

RA Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;

RT "Analysis of the Escherichia coli genome: DNA sequence of the region
 RT from 84.5 to 86.5 minutes.";

RL Science 257:771-778(1992).

RN 13

RP REVISION TO 14-15.

RC STRAIN=K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.D.,
 RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1453-1474(1997).

RN 14

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postell G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobbeck E.J., Davis N.W., Lim A., Diallanta E.T., Potomous K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;

RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";

RL Nature 409:529-533(2001).

RN 15

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / RMD 0509952;

RX MEDLINE=21156231; PubMed=11258796;

RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han G.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kubara S., Shiba T., Hattori M., Shingawa H.;

RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";

RL DNA Res. 8:11-22(2001).

RN 16

RP SEQUENCE OF 30-289 FROM N.A.

RC STRAIN=K12;

RX MEDLINE=85003590; PubMed=6383820;

RA de Geus P., Verheij H.M., Riegmans N.H., Hoeksma W.P.M., de Haas G.H.;

RT "The pro- and mature forms of the E. coli K-12 outer membrane
 RT phospholipase A are identical.";

RL EMBO J. 3:1799-1802(1984).

RN 17

RP SEQUENCE OF 174-289 FROM N.A.

RC STRAIN=K12;

RX MEDLINE=87115164; PubMed=3027506;

RA Irino N., Nakayama K., Nakayama H.;

RT "The recG gene of Escherichia coli K12: primary structure and
 RT evidence for SOS regulation.";

RL Mol. Gen. Genet. 205:298-304(1986).

RN 18

RP MUTAGENESIS OF SER-172.

RX MEDLINE=94131966; PubMed=8300539;

RA Brok R.G.P.M., Brinkman E., van Boxel R., Bekkers A.C.A.P.,
 RA Verheij H.M., Tomassen J.;

RT "Molecular characterization of enterobacterial plid genes encoding
 RT outer membrane phospholipase A.";

RL J. Bacteriol. 176:861-870(1994).

RN 19

RP ACTIVE SITE SER-164, AND PARTIAL SEQUENCE.

RX MEDLINE=91249806; PubMed=2040286;

RA Horrovoets A.J.G., Verheij H.M., de Haas G.H.;

RT "Inactivation of Escherichia coli outer-membrane phospholipase A by
 RT the affinity label hexadecanesulfonyl fluoride. Evidence for an
 RT active-site serine.";

RL Eur. J. Biochem. 198:247-253(1991).

CC -1- FUNCTION: HYDROLYSIS OF PHOSPHATIDYLCHOLINE WITH PHOSPHOLIPASE
 CC A2 (EC 3.1.1.4) AND PHOSPHOLIPASE A1 (EC 3.1.1.32) ACTIVITIES.
 CC REQUIRED FOR EFFICIENT SECRETION OF BACTERIOCIDINS, SEEMS TO BE
 CC DORMANT IN NORMAL GROWING CELLS.

CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 1-ACYLGLYCERO-
 CC PHOSPHOCHOLINE + A FATTY ACID ANION.

CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 2-ACYLGLYCERO-
 CC PHOSPHOCHOLINE + A FATTY ACID ANION.

CC -1- COFACTOR: REQUIRES CALCIUM IONS FOR ACTIVITY.

CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE; ONE OF THE VERY FEW ENZYMES
 CC LOCATED THERE.

CC -1- INDUCTION: BY MEMBRANE DAMAGING, FOR EXAMPLE, BY PHAGE-INDUCED
 CC LYSIS OR TEMPERATURE SHOCK.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: X02143; CAA26081.1; -
 DR EMBL: M87049; AA67617.1; -
 DR EMBL: AE000458; AAC76824.1; -
 DR EMBL: AE005613; AAC59017.1; -
 DR EMBL: AP002567; BAB38174.1; -
 DR EMBL: M30198; AAA24516.1; -
 DR PIR: A07711; PSECA.
 DR PIR: A22133; PSECA1.
 DR PIR: S30711; S30711.
 DR Ecocore: EG10738; PLDA.
 DR InterPro: IPR003187; PLAL.
 DR Pfam: PF02253; PLAL; 1.
 KW Hydrolase; Lipid degradation; Outer membrane; Signal; Calcium;
 KW Complete proteome.
 FT SIGNAL 1 20
 FT CHAIN 21 289 PHOSPHOLIPASE A1.
 FT ACT_SITE 164 164
 FT MUTAGEN 172 172 S->F: INACTIVE PROTEIN.
 FT CONFLICT 14 15 LP -> FA (IN REF. 2).
 FT CONFLICT 30 33 DAPA -> MTRQ (IN REF. 6).
 SQ SEQUENCE 289 AA; 33163 MW; A688AD32AA60F218 CRC64;

Query Match 93.3%; Score 28; DB 1; Length 289;
 Best Local Similarity 80.0%; Pred. No. 97;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 1111
 DB 95 FPLMR 99

Search completed: February 27, 2002, 11:42:48
 Job time: 547 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:50:14 : Search time 281.76 seconds
(Without alignments)
2.596 Million cell updates/sec

Title: US-09-446-109a-16
Perfect score: 30
Sequence: 1 FPMXR 5

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP RODENT:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	96.7	101	5	09V671 drosophila
2	29	96.7	113	10	039293 brassica na
3	29	96.7	157	5	09X7Y5 09x7y5 caenorhabd
4	29	96.7	158	4	09NNW3 09nnw3 homo sapien
5	29	96.7	158	4	09NNW2 09nnw2 homo sapien
6	29	96.7	229	8	037633 thagolelis
7	29	96.7	245	2	09PDE1 09pdel xyella fas
8	29	96.7	273	10	09SUV9 09suv9 arabidopsi
9	29	96.7	292	2	09HT83 09ht83 pseudomonas
10	29	96.7	321	11	09D6L7 09d6l7 mus musculu
11	29	96.7	332	2	09APJ5 09apj5 erwinia chr
12	29	96.7	333	2	09LJCG 09ljcg erwinia rha
13	29	96.7	334	10	09FTM4 09ftm4 oryza sativ
14	29	96.7	349	4	015429 homo sapien
15	29	96.7	417	2	09SMA1 09sma1 arabidopsi
16	29	96.7	456	2	09PDK4 09pdk4 zymomonas m
17	29	96.7	479	2	086742 streptomyces
18	29	96.7	497	2	091702 pseudomonas
19	29	96.7	501	5	016923 caenorhabd

20	29	96.7	502	5	09GCU15 09gcu15 caenorhabd
21	29	96.7	555	1	029748 029748 archaeoglob
22	29	96.7	581	4	09BR01 09br01 homo sapien
23	29	96.7	612	4	09HRY2 09hry2 homo sapien
24	29	96.7	621	4	09H9Y1 09h9y1 homo sapien
25	29	96.7	625	2	066576 066576 aquifex aeo
26	29	96.7	634	2	09NNW1 09nnw1 homo sapien
27	29	96.7	634	4	09NWX5 09nwx5 homo sapien
28	29	96.7	642	2	025396 025396 helicobacte
29	29	96.7	642	2	09ZLF3 09zlf3 helicobacte
30	29	96.7	650	5	017866 017866 caenorhabd
31	29	96.7	735	2	09HU63 09hu63 pseudomonas
32	29	96.7	778	5	09U9K6 09u9k6 caenorhabd
33	29	96.7	811	4	09BWX2 09bwx2 homo sapien
34	29	96.7	1071	11	09D2K4 09d2k4 mus musculu
35	29	96.7	1108	13	09PMD0 09pmd0 tetraodon f
36	29	96.7	1117	5	09U9K7 09u9k7 caenorhabd
37	29	96.7	1451	2	045336 045336 bordetella
38	29	96.7	1451	2	045044 045044 bordetella
39	29	96.7	1464	2	09S5D5 09s5d5 bordetella
40	28	93.3	68	2	P73858 p73858 synechocyst
41	28	93.3	84	2	P74471 p74471 synechocyst
42	28	93.3	84	2	P73028 p73028 synechocyst
43	28	93.3	93	2	055568 055568 synechocyst
44	28	93.3	103	10	09M045 09m045 arabidopsi
45	28	93.3	110	6	09GLJ9 09glj9 canis faml

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	101 AA.
Q9V671	Q9V671			
AC	Q9V671			
BT	01-MAY-2000 (TREMURel. 13, Created)			
LT	01-MAY-2000 (TREMURel. 13, Last sequence update)			
DT	01-JUN-2001 (TREMURel. 17, Last annotation update)			
DE	CG13174 PROTEIN.			
CN	CG13174.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
CC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
CC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
EA	[1]			
EA	SEQUENCE FROM N.A.			
RP	STRAIN=BERKELEY;			
RC	MEDLINE=20196006; PubMed=10731132;			
KA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
KA	Amannatides P.C., Scherfer S.E., Li P.W., Hoskins R.A., Galle R.F.,			
KA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
KA	Sulton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,			
KA	Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Peiffer B.D.,			
KA	Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,			
KA	April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,			
KA	Bailey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,			
KA	Beeson K.M., Benos P.V., Berman B.P., Bhandari D., Boltskov S.,			
KA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,			
KA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,			
KA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,			
KA	de Pablos B., Delcher A., Deng Z., Dey A.D., Dew I., Dietz S.M.,			
KA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,			
KA	Durbin K.J., Evangelista C.C., Ferrez C., Fertala S., Fleischmann W.,			
KA	Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,			
KA	Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,			
KA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,			
KA	Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Jengam C.,			
KA	Jatelli M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,			
KA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,			
KA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,			
KA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,			
KA	Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,			

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
 RA Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Schaefer F., Shen H.,
 RA Shue B.C., Sinden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Stryckas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Morley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2183-2193(2000).
 DR EMBL: AE003823; AAF58560.1; -;
 DR FlyBase: FBgn003694; CG13174.
 SO SEQUENCE 101 AA; 11820 MW; D5C4D85B514CB9B CRC64;

Query Match 96.7%: Score 29; DB 5; Length 101;
 Best Local Similarity 80.0%: Pred. No. 1.2e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FPMXR 5
 DB 75 FPMXR 79

RESULT 2
 ID 039293 PRELIMINARY; PRT; 113 AA.
 AC 039293;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DE ELONGATION FACTOR EF-1A (FRAGMENT).
 OS Brassica napus (Rape).
 OC Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Brassica.
 OX NCBI_TaxID=3708;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SAMOURAI; TISSUE=ETIOLATED SEEDLINGS, (5-DAYS OLD);
 RA Saez-Vasquez J.;
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO GTP-BINDING ELONGATION FACTOR FAMILY.
 DR EMBL: U21744; AAA86366.1; -;
 DR Mendel: 15767; Brana:1139;15767.
 DR InterPro: IPR000795; GTP_EFTU.
 DR Pfam: PF00009; GTP_EFTU; 1.
 KW Elongation factor; GTP-binding; Protein biosynthesis.
 FT NON_TER 1
 SO SEQUENCE 113 AA; 12560 MW; 149783D707223948 CRC64;

Query Match 96.7%: Score 29; DB 10; Length 113;
 Best Local Similarity 80.0%: Pred. No. 1.4e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FPMXR 5
 DB 103 FPMXR 107

RESULT 3
 O9XTY5 PRELIMINARY; PRT; 157 AA.
 AC O9XTY5;
 DT 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE T03D8.2 PROTEIN.

CN T03D8.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Mortimore B.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=4150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierley-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstein L., Wilkinson-Sproat J., Wohlman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*."
 RL Nature 368:32-38(1994).
 DR EMBL: Z92838; CAB07406.1; -;
 DR InterPro: IPR000230; Ribosomal_S12.
 DR Pfam: PF00164; Ribosomal_S12; 1.
 DR PRINTS: PR01034; RIBOSOMALS12.
 DR Prodom: PD000576; Ribosomal_S12; 1.
 SO SEQUENCE 157 AA; 17370 MW; 6C6A837282EB26C CRC64;

Query Match 96.7%: Score 29; DB 5; Length 157;
 Best Local Similarity 80.0%: Pred. No. 1.9e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FPMXR 5
 DB 40 FPMXR 44

RESULT 4
 ID Q9NNW3 PRELIMINARY; PRT; 158 AA.
 AC Q9NNW3;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE MEMBRANE COFACTOR PROTEIN CD46 VARIANT (FRAGMENT).
 GN MCP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20227770; PubMed=10751138;
 RA Kusuhara K., Sasaki Y., Nakao F., Ihara K., Hattori H., Yamashita S.,
 RA Nihei K., Koide N., Alba H., Takeshita K., Hara T.;
 RT "Analysis of measles virus binding sites of the CD46 gene in patients
 with subacute sclerosing panencephalitis."
 RL J. Infect. Dis. 181:1447-1449(2000).
 DR EMBL: AF209713; AAF73845.1; -;
 DR InterPro: IPR000436; Sush1_SCR_CCP.
 DR Pfam: PF00084; Sush1; 2.
 DR SMART: SM00032; CCP; 2.
 DR VARIANT 59
 FT NON_TER 158
 SO SEQUENCE 158 AA; 18098 MW; 4E3F07EBC5C454F5 CRC64;

Query Match 96.7%: Score 29; DB 4; Length 158;

Best Local Similarity 80.0% Pred. No. 1.9e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FPMR 5
11 11
Db 11 FPMR 15

RESULT 5
ID 09NNM2 PRELIMINARY; PRT; 158 AA.
AC 09NNM2:
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE MEMBRANE COFACTOR PROTEIN CD46 VARIANT (FRAGMENT).
CN MCP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20227770; PubMed=10751138;
RA Kusuhara K., Sasaki Y., Nakao F., Ihara K., Hattori H., Yamashita S.,
RA Mihel K., Kolde N., Alba H., Takeshita K., Hara T.;
RT "Analysis of measles virus binding sites of the CD46 gene in patients
with subacute sclerosing panencephalitis."
RL J. Infect. Dis. 181:1447-1449(2000).
DR EMBL; AF209714; AAF73846.1;
DR InterPro: IPR000436; Sush1_SCR_CCP.
DR Pfam: PF00084; sush1; 2.
DR SMART: SM00032; CCP; 2.
FT NON_TER 158 158
SQ SEQUENCE 158 AA: 18126 MW: 82CF38EC3A16A9DA CRC64:

Query Match 96.7% Score 29; DB 4; Length 158;
Best Local Similarity 80.0% Pred. No. 1.9e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FPMR 5
11 11
Db 11 FPMR 15

RESULT 6
ID 037633 PRELIMINARY; PRT; 229 AA.
AC 037633:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1) (FRAGMENT).
CN COI.
OS Rhagoletis boycel.
OC Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Tephritidae; Tephritidae; Rhagoletis.
OX NCBI_TaxID=43419;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97159559; PubMed=9007018;
RA Smith J.J., Bush G.L.;
RT "Phylogeny of the genus Rhagoletis (Diptera: Tephritidae) inferred
from DNA sequences of mitochondrial cytochrome oxidase II."
RL Mol. Phylogenet. Evol. 7:33-43(1997).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 2
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 1-
CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
CC A CENTER TO THE BINUCLEAR CENTER OF THE CATALYTIC SUBUNIT 1 (BY

CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O + 4
CC FERRICYTOCHROME C.
CC -1- COFACTOR: COPPER A AND HEME GROUP.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE (BY SIMILARITY).
CC -1- SIMILARITY: TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.

DR EMBL; U53254; AAB50355.1;
DR HSSP; P08306; IAR1.
DR InterPro: IPR001505; COX2.
DR InterPro: IPR002429; Cyt_c-ox_2.
DR Pfam: PF00116; COX2; 1.
DR PRINTS; PR01166; CYCOXIDASEII.
DR PRODOM; PD000131; COX2; 1.
DR PROSITE; PS00078; COX2; 1.
KW Copper; Inner membrane; Mitochondrion; Oxidoreductase;
KW Respiratory chain; Transmembrane.
FT METAL 159 159 COPPER A (PROBABLE).
FT METAL 163 163 COPPER A (PROBABLE).
FT METAL 167 167 COPPER A (PROBABLE).
FT METAL 207 207 COPPER A (PROBABLE).
FT NON_TER 229 229
SQ SEQUENCE 229 AA: 26389 MW: 912D1B2736330698 CRC64:

Query Match 96.7% Score 29; DB 8; Length 229;
Best Local Similarity 80.0% Pred. No. 2.7e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FPMR 5
11 11
Db 78 FPMR 82

RESULT 7
ID 09PDEL PRELIMINARY; PRT; 245 AA.
AC 09PDEL:
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE HYPOTHETICAL PROTEIN XP1438.
CN XP1438.
OS Xylella fastidiosa.
CC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorfi H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
RA Gantier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohnselt J.D., Junqueira M.L., Kemper E.L., Kitaajima J.P.,
RA Krieger J.E., Kuramae E.L., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Martino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A., Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Pelxoto B.R., Pereira G.A.G., Pereira H.A., Jr., Pesquero J.B.,
RA Quagiro R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E., Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A., Jr.,

RA de Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsunako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*,";
 RL Nature 406:151-159(2000).
 DR EMBL: AE003974; AAF84247.1; -
 KW Hypothetical protein; Complete proteome.
 SO SEQUENCE 245 AA; 27193 MW; 16C8ADB9FC490455 CRC64;

Query Match: 96.7%; Score 29; DB 2; Length 245;
 Best Local Similarity 80.0%; Pred. NO. 2.8e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 Db 197 FPSWR 201

RESULT 8
 ID 09SUV9 PRELIMINARY; PRT; 273 AA.
 AC 09SUV9;
 DT 01-MAY-2000 (TREMUREL. 13, Created)
 DT 01-MAY-2000 (TREMUREL. 13, Last sequence update)
 DE 01-OCT-2000 (TREMUREL. 15, Last annotation update)
 DE HYPOTHETICAL 31.2 KDA PROTEIN.
 CN FTK2.180 OR AT4G22600.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Medler H., Wambuit R., Bancroft I., Mewes H.W.,
 RA Mayer K.F.X., Lemcke K., Schueller G.;
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Medler H., Wambuit R., Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL033545; CAA22165.1; -
 DR EMBL: AL161557; CAB79215.1; -
 KW Hypothetical protein.
 SO SEQUENCE 273 AA; 31248 MW; 7E644F686CE8EBB0 CRC64;

Query Match: 96.7%; Score 29; DB 10; Length 273;
 Best Local Similarity 80.0%; Pred. NO. 3.1e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 Db 79 FPSWR 83

RESULT 9
 ID 09HTB3 PRELIMINARY; PRT; 292 AA.
 AC 09HTB3;
 DT 01-MAR-2001 (TREMUREL. 16, Created)
 DT 01-MAR-2001 (TREMUREL. 16, Last sequence update)
 DT 01-JUN-2001 (TREMUREL. 17, Last annotation update)

DE HYPOTHETICAL PROTEIN PA5488.
 GN PA5488.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PA01;
 RX MEDLINE=20437337; Pubmed=10984043;
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle M.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Gentry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olsson M.V.;
 RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an
 opportunistic pathogen,";
 XL Nature 406:959-964(2000).
 DR EMBL: AE004361; AAG08873.1; -
 DR Interpro: IPR000504; RRM.
 DR PROSITE: PS00030; RRM_RNP_1; UNKNOWN_1.
 KW Hypothetical protein; Complete proteome.
 SO SEQUENCE 292 AA; 32949 MW; 14610337E8CCDFD4 CRC64;

Query Match: 96.7%; Score 29; DB 2; Length 292;
 Best Local Similarity 80.0%; Pred. NO. 3.1e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 Db 231 FPSWR 235

RESULT 10
 ID 09D6L7 PRELIMINARY; PRT; 321 AA.
 AC 09D6L7;
 DT 01-JUN-2001 (TREMUREL. 17, Created)
 DT 01-JUN-2001 (TREMUREL. 17, Last sequence update)
 DT 01-JUN-2001 (TREMUREL. 17, Last annotation update)
 DE 2310076014RIK PROTEIN.
 GN 2310076014RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=KIDNEY;
 RX MEDLINE=21085660; Pubmed=11217851;
 RA Kawai J., Shimagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staudli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsi G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitlaker C., Wilming L.,
 RA Wysshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection,";
 XL Nature 409:685-690(2001).
 DR EMBL: AK010208; BAB26769.1; -

DR MCD: MGI:1919189; 231007601ARik.
 DR InterPro: IPR000847; HTH_LYSR..
 DR PROSITE: PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.
 SQ SEQUENCE 321 AA; 35038 MW; 6E27EBC746552DB8 CRC64;

Query Match
 Best Local Similarity 96.7%; Score 29; DB 11; Length 321;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 DB 44 FPSMR 48

RESULT 11

ID O9APJ5 PRELIMINARY; PRT: 332 AA.
 AC O9APJ5;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE ENDO-1,4-BETA-D-GLUCANASE PRECURSOR.
 GN CELBY.
 OS Erwinia chrysanthemi.
 OC Bacteria; Proteobacteria; gamma subdivision: Enterobacteriaceae;
 OC Pectobacterium.
 OX NCBI_TaxID=556;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN=PY35;
 RA Cho S.J., Park S.R., Yun H.D.;
 RT "Cloning and sequencing of a celby gene of Pectobacterium chrysanthemi
 RT PY35."
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF282321; AAC49556.1; -.
 KW SIGNAL.
 FT SIGNAL 1 23 POTENTIAL.
 SQ SEQUENCE 332 AA; 37627 MW; A1B5D7B0CB820EE6 CRC64;

Query Match
 Best Local Similarity 96.7%; Score 29; DB 2; Length 332;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 DB 178 FPAWR 182

RESULT 12

ID O9LJG9 PRELIMINARY; PRT: 333 AA.
 AC O9LJG9;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4).
 GN CELA.
 OS Erwinia rhapsodict.
 OC Bacteria; Proteobacteria; gamma subdivision: Enterobacteriaceae;
 OC Erwinia.
 OX NCBI_TaxID=55212;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCPPB2989;
 RA Saarikallio H.T.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN 121
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCPPB2989;
 RA Riekl R.;
 OS "Members of the amylovora group of Erwinia are cellulolytic and

RT possess genes homologous to the type II secretion pathway."
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ276358; CAB89803.1; -.
 DR InterPro: IPR002037; Glyco_hydro_8.
 DR Pfam: PF01270; Glyco_hydro_8; 1.
 DR PRINTS: PR00735; GLYDRLASER.
 KW Signal: Hydrolase; Glycosidase.
 FT SIGNAL 1 23
 FT CHAIN 24 333
 SQ SEQUENCE 333 AA; 37783 MW; D75CFEF212302673A CRC64;

Query Match
 Best Local Similarity 96.7%; Score 29; DB 2; Length 333;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 DB 178 FPAWR 182

RESULT 13

ID O9FTM4 PRELIMINARY; PRT: 334 AA.
 AC O9FTM4;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE P0005A05.25 PROTEIN.
 GN P0005A05.25.
 OS Oryza sativa (Rice).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 CC Ehrhartoideae; Oryzaceae; Oryza.
 GN NCBI_TaxID=4540;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 RT clone:P0005A05."
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AP002863; BAB16921.1; -.
 DR InterPro: IPR000520; Exonuclease.
 DR InterPro: IPR000822; Znf-C2H2.
 DR Pfam: PF000726; Exonuclease; 1.
 DR SMART: SM01 E-C2H2; 1.
 DR SMART: SM01 EXONUC; 1.
 DR PROSITE: PROSITE_2; 2.
 DR PROSITE: PROSITE_1; 1.
 KW DNA-binding
 SQ SEQUENCE 334 AA; 37172 MW; A34531D625704A88 CRC64;

Query Match
 Best Local Similarity 96.7%; Score 29; DB 10; Length 334;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 DB 301 FPAWR 305

RESULT 14

ID O15429 PRELIMINARY; PRT: 349 AA.
 AC O15429;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE CD46.
 OS Homo sapiens (Human).

Search completed: February 27, 2002, 11:50:14
 Job time: 993 sec

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=TESTIS;
 RA Hara T., Suzuki Y., Nagasawa S., Seya T.,
 RT "CDNA cloning and characterization of human sperm CD46."
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: D84105; BAA12224.1; -
 DR HSSP: P10998; IVD.
 DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF00084; Sushi; 4.
 DR SMART: SM00032; CCP; 4.
 SO SEQUENCE 349 AA; 39325 MW; 8EFCEDA30D3C818E CRC64;

Query Match 96.7% Score 29; DB 4; Length 349;
 Best Local Similarity 80.0% Pred. No. 4e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
 11 11
 Db 11 FPXWR 15

RESULT 15
 O9SAA1
 ID O9SAA1 PRELIMINARY; PRT: 417 AA.
 AC O9SAA1:
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE F25C20.9.
 DE F25C20.9.
 GN F25C20.9.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Vysotskaya V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
 RA Li J., Lee J.M., Kremetskaya I., Luros J., Ngan I., Liu A.,
 RA Gonzalez A., Altafi H., Araujo R., Chao O., Conn L., Conway A.B.,
 RA Dunn P., Hansen N., Huizar L., Kim C., Palm C., Rowley D., Shin P.,
 RA Walker M., Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.,
 RT "Arabidopsis thaliana chromosome 1 BAC F25C20 sequence."
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Theologis A.,
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Theologis;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC007296; AAC0247.1; -
 SO SEQUENCE 417 AA; 45985 MW; 886EBF06AC19A63D CRC64;

Query Match 96.7% Score 29; DB 10; Length 417;
 Best Local Similarity 80.0% Pred. No. 4.7e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
 11 11
 Db 377 FPXWR 381

200

THE UNIVERSITY OF CHICAGO
LIBRARY

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:36:07 ; Search time 132.19 seconds
(without alignments)
0.851 Million cell updates/sec

Title: US-09-446-109A-16

Sequence: 1 FPXMR 5

Scoring table: BLOSUM62
Gapop 10.0, Capext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database:

Issued Patents_AA:*

- 1: /cgn2_6/prodata2/1aa/5A_COMB.pep:*
- 2: /cgn2_6/prodata2/1aa/5B_COMB.pep:*
- 3: /cgn2_6/prodata2/1aa/5A_COMB.pep:*
- 4: /cgn2_6/prodata2/1aa/5B_COMB.pep:*
- 5: /cgn2_6/prodata2/1aa/PCTUS_COMB.pep:*
- 6: /cgn2_6/prodata2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	96.7	38	1 US-08-176-500-118	Sequence 118, App
2	29	96.7	38	1 US-08-471-052A-118	Sequence 118, App
3	29	96.7	38	1 US-08-189-331-118	Sequence 118, App
4	29	96.7	38	2 US-08-471-939-118	Sequence 118, App
5	29	96.7	38	2 US-08-471-800-118	Sequence 118, App
6	29	96.7	38	2 US-08-471-068-118	Sequence 118, App
7	29	96.7	241	4 US-08-823-120-1	Sequence 11, Appl
8	29	96.7	324	2 US-08-528-057-46	Sequence 46, Appl
9	29	96.7	370	2 US-08-528-057-42	Sequence 42, Appl
10	29	96.7	372	1 US-08-202-056-5	Sequence 5, Appl
11	29	96.7	372	1 US-08-076-093A-6	Sequence 6, Appl
12	29	96.7	372	1 US-08-701-265-6	Sequence 6, Appl
13	29	96.7	372	2 US-08-284-586-6	Sequence 6, Appl
14	29	96.7	372	2 US-08-805-478-6	Sequence 6, Appl
15	29	96.7	372	2 US-08-802-627A-6	Sequence 6, Appl
16	29	96.7	372	2 US-08-801-238-6	Sequence 6, Appl
17	29	96.7	372	2 US-08-801-228-6	Sequence 6, Appl
18	29	96.7	372	3 US-09-104-296-6	Sequence 6, Appl
19	29	96.7	372	3 US-08-982-493-8	Sequence 8, Appl
20	29	96.7	373	2 US-08-528-057-44	Sequence 44, Appl
21	29	96.7	377	2 US-08-528-057-2	Sequence 2, Appl
22	29	96.7	384	4 US-08-139-195-2	Sequence 2, Appl
23	29	96.7	384	6 5514787-2	Patent No. 5514787
24	28	93.3	57	1 US-08-370-225-29	Sequence 29, Appl
25	28	93.3	57	1 US-08-370-225-30	Sequence 30, Appl
26	28	93.3	57	1 US-08-461-859-29	Sequence 29, Appl
27	28	93.3	57	1 US-08-461-859-30	Sequence 30, Appl

28	28	93.3	57	5 PCT-US93-10069-29	Sequence 29, Appl
29	28	93.3	57	5 PCT-US93-10069-30	Sequence 30, Appl
30	28	93.3	298	2 US-08-061-636-3	Sequence 3, Appl
31	28	93.3	298	2 US-08-874-347-19	Sequence 19, Appl
32	28	93.3	298	5 US-09-093-522-19	Sequence 19, Appl
33	28	93.3	298	5 PCT-US94-05268-3	Sequence 3, Appl
34	28	93.3	373	2 US-08-846-762-13	Sequence 13, Appl
35	28	93.3	374	3 US-08-982-493-6	Sequence 6, Appl
36	28	93.3	437	2 US-09-031-059-1	Sequence 1, Appl
37	28	93.3	437	2 US-09-031-059-3	Sequence 3, Appl
38	28	93.3	760	1 US-08-195-152-2	Sequence 2, Appl
39	26	86.7	10	4 US-09-461-697-409	Sequence 409, App
40	26	86.7	12	4 US-09-461-697-407	Sequence 407, App
41	26	86.7	113	1 US-07-668-648-10	Sequence 10, App
42	26	86.7	113	2 US-08-429-998-10	Sequence 10, App
43	26	86.7	113	2 US-08-431-333-10	Sequence 10, App
44	26	86.7	113	5 PCT-US91-02321-10	Sequence 10, App
45	26	86.7	862	1 US-08-325-267A-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-176-500-118
Sequence 118, Application US/08176500
Patent No. 5498538
GENERAL INFORMATION:
APPLICANT: Kay, R. K.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 141
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/176,500
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/013,416
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-143
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-176-500-118

Query Match 96.7%; Score 29; DB 1; Length 38;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
DB 10 FPSMR 14

RESULT 2

US-08-471-052A-118
Sequence 118, Application US/08471052A
Patent No. 5625033
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 166
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,052A
FILING DATE: 06-JUNE-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mistock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-179
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-471-052A-118

Query Match 96.7%; Score 29; DB 1; Length 38;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
DB 10 FPSMR 14

RESULT 3

US-08-189-331-118
Sequence 118, Application US/08189331
Patent No. 5747334
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 186
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/189,331
FILING DATE: Concurrently herewith
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mistock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-155
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-189-331-118

Query Match 96.7%; Score 29; DB 1; Length 38;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
DB 10 FPSMR 14

RESULT 4

US-08-471-939-118
Sequence 118, Application US/08471939
Patent No. 5844076
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 141
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,939
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/013,416
FILING DATE: 01-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mistock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-143
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:

SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-471-939-118

Query Match 96.7%: Score 29; DB 2; Length 38;
Best Local Similarity 80.0%: Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
DB 10 FPSMR 14

RESULT 5
US-08-471-800-118
Sequence 118, Application US/08471800
Patent No. 5852167
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 141
CORRESPONDENCE ADDRESS:
ADDRESSEE: Penile & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,800
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/013,416
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mistock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-143
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-471-800-118

Query Match 96.7%: Score 29; DB 2; Length 38;
Best Local Similarity 80.0%: Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
DB 10 FPSMR 14

RESULT 6
US-08-471-068-118
Sequence 118, Application US/08471068
Patent No. 5948635
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 186
CORRESPONDENCE ADDRESS:
ADDRESSEE: Penile & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,068
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/189,331
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mistock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-155
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-471-068-118

Query Match 96.7%: Score 29; DB 2; Length 38;
Best Local Similarity 80.0%: Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
DB 10 FPSMR 14

RESULT 7
US-08-823-120-1
Sequence 1, Application US/08823120
Patent No. 6148919
GENERAL INFORMATION:
APPLICANT: Domenighini, Mario
APPLICANT: Rappuoli, Rino
TITLE OF INVENTION: Immunogenic Detoxified Mutants of
TITLE OF INVENTION: Cholera toxin and of the toxin lt. Their preparation and
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: California
COUNTRY: USA

ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/823,120
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/256,003
FILING DATE: 11-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: McCLUNG, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0315,001
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 241 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-823-120-1

Query Match 96.7%; Score 29; DB 4; Length 241;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
db 169 FPMR 173

RESULT 8
US-08-528-057-46
Sequence 46, Application US/08528057
Patent No. 5846715
GENERAL INFORMATION:
APPLICANT: PURCELL, Damian F. J.
APPLICANT: RUSSELL, Sarah M.
APPLICANT: MCKENZIE, Ian F. C.
TITLE OF INVENTION: CD46 VARIANTS
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
City: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/528,057
FILING DATE: CONCURRENTLY HEREWITH
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/961,686
FILING DATE: 11-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU91/00199
FILING DATE: 10-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PK0133/90
FILING DATE: 11-MAY-1990

ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 17227/112 DACO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 324 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-528-057-46

Query Match 96.7%; Score 29; DB 2; Length 324;
Best Local Similarity 80.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
db 11 FPMR 15

RESULT 9
US-08-528-057-42
Sequence 42, Application US/08528057
Patent No. 5846715
GENERAL INFORMATION:
APPLICANT: PURCELL, Damian F. J.
APPLICANT: RUSSELL, Sarah M.
APPLICANT: MCKENZIE, Ian F. C.
TITLE OF INVENTION: CD46 VARIANTS
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
City: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/528,057
FILING DATE: CONCURRENTLY HEREWITH
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/961,686
FILING DATE: 11-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU91/00199
FILING DATE: 10-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PK0133/90
FILING DATE: 11-MAY-1990
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 17227/112 DACO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 370 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-528-057-42

Query Match 96.7%; Score 29; DB 2; Length 370;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPKWR 5
11 11
Db 11 FPKWR 15

RESULT 10
US-08-202-056-5
Sequence 5, Application US/08202056
Patent No. 5440021
GENERAL INFORMATION:
APPLICANT: Chuntharapal, Anan
APPLICANT: Hebert, Caroline
APPLICANT: Kim, Kyung Jin
APPLICANT: Lee, James
TITLE OF INVENTION: Antibodies to Human IL-8 Type B Receptor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: palin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202,056
FILING DATE: 25-FEB-1994
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/677211
FILING DATE: 29-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B.
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: 706P3
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-202-056-5

Query Match 96.7%; Score 29; DB 1; Length 372;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPKWR 5
11 11
Db 352 FPKWR 356

RESULT 11
US-08-076-093A-6
Sequence 6, Application US/08076093A
Patent No. 5543503
GENERAL INFORMATION:

APPLICANT: Chuntharapal, Anan
APPLICANT: Lee, James
APPLICANT: Hebert, Caroline
APPLICANT: Jin Kim, K.
TITLE OF INVENTION: Antibodies to Human PFAA Receptors
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

QY 1 FPKWR 5
11 11
Db 352 FPKWR 356

RESULT 12
US-08-701-265-6
Sequence 6, Application US/08701265
Patent No. 577,17
GENERAL INFORMATION:
APPLICANT: Chuntharapal, Anan
APPLICANT: Lee, James
APPLICANT: Hebert, Caroline
APPLICANT: Jin Kim, K.
TITLE OF INVENTION: Antibodies to Human PFAA Receptors
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

Query Match 96.7%; Score 29; DB 1; Length 372;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0

US-08-076-093A-6

COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/701,265
FILING DATE: 22-AUG-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/076093
FILING DATE: 11-JUN-1993
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
PRIOR APPLICATION DATA: 07/677211
FILING DATE: 29-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: 706P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-701-265-6

Query Match 96.7%; Score 29; DB 1; Length 372;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXWR 5
11 11
DB 352 FPSWR 356

RESULT 13
US-08-284-586-6
Sequence 6, Application US/08284586
Patent No. 5840856
GENERAL INFORMATION:
APPLICANT: Chuntharapal, Anan
APPLICANT: Lee, James
APPLICANT: Hebert, Caroline
APPLICANT: Jin Kim, K.
TITLE OF INVENTION: Antibodies to Human PFAA Receptors
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/284,586
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/076,093A
FILING DATE: 11-JUN-1993
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/677211
FILING DATE: 29-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: 706P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-284-586-6

Query Match 96.7%; Score 29; DB 2; Length 372;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXWR 5
11 11
DB 352 FPSWR 356

RESULT 14
US-08-805-478-6
Sequence 6, Application US/08805478
Patent No. 5874543
GENERAL INFORMATION:
APPLICANT: Chuntharapal, Anan
APPLICANT: Lee, James
APPLICANT: Hebert, Caroline
APPLICANT: Jin Kim, K.
TITLE OF INVENTION: ANTIBODIES TO PFAA RECEPTOR
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/805,478
FILING DATE: 25-FEB-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/284586
FILING DATE: 10-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/076093
FILING DATE: 11-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B.
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: P0706P2P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:

LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-805-478-6

Search completed: February 27, 2002, 11:36:07
Job time: 146 sec

Query Match 96.7%; Score 29; DB 2; Length 372;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
DB 352 FPSWR 356

RESULT 15
US-08-802-627A-6
Sequence 6, Application US/08802627A
Patent No. 5892017
GENERAL INFORMATION:
APPLICANT: Lee, James
APPLICANT: Wood, William L.
TITLE OF INVENTION: NUCLEIC ACID ENCODING PFAA RECEPTOR
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPacIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/802,627A
FILING DATE: 19-Feb-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/284586
FILING DATE: 10-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/076093
FILING DATE: 11-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B.
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: P0706P2PID2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-802-627A-6

Query Match 96.7%; Score 29; DB 2; Length 372;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
DB 352 FPSWR 356

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:41:19 : Search time 303.5 Seconds

(without alignments)
1.464 Million cell updates/sec

Title: us-09-446-109a-17

Perfect score: 31

Sequence: 1 FXPWMR 6

Scoring table: BLOSUM62

Gapop 10.0, Capext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq_1101.*
1: /SID8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	90.3	71	22	AA16826
2	28	90.3	71	22	AA16826
3	28	90.3	71	22	AA16826
4	28	90.3	277	21	AA16826
5	28	90.3	357	21	AA16826
6	28	90.3	383	22	AA16826
7	28	90.3	488	20	AA16826
8	28	90.3	509	22	AA16826
9	28	90.3	509	22	AA16826
10	28	90.3	527	19	AA16826
11	28	90.3	527	19	AA16826

12	28	90.3	629	20	AA16826
13	28	90.3	766	22	AA16826
14	28	90.3	949	22	AA16826
15	28	90.3	2247	18	AA16826
16	27	87.1	6	13	AA16826
17	27	87.1	6	16	AA16826
18	27	87.1	30	22	AA16826
19	27	87.1	31	22	AA16826
20	27	87.1	39	22	AA16826
21	27	87.1	79	22	AA16826
22	27	87.1	312	22	AA16826
23	27	87.1	359	22	AA16826
24	27	87.1	392	21	AA16826
25	27	87.1	399	21	AA16826
26	27	87.1	408	21	AA16826
27	27	87.1	434	21	AA16826
28	27	87.1	434	21	AA16826
29	27	87.1	449	20	AA16826
30	27	87.1	449	20	AA16826
31	27	87.1	499	21	AA16826
32	27	87.1	588	21	AA16826
33	27	87.1	594	21	AA16826
34	27	87.1	704	22	AA16826
35	27	87.1	719	22	AA16826
36	27	87.1	724	22	AA16826
37	27	87.1	761	20	AA16826
38	27	87.1	761	21	AA16826
39	27	87.1	846	21	AA16826
40	27	87.1	897	22	AA16826
41	27	87.1	1093	14	AA16826
42	27	87.1	1204	21	AA16826
43	26	83.9	160	22	AA16826
44	26	83.9	254	19	AA16826
45	26	83.9	262	21	AA16826

ALIGNMENTS

RESULT 1
ID AA16826 standard; Protein: 71 AA.
AC AA16826;
XX
DT 12-OCT-2001 (first entry)
XX
ID Peptide #3260 encoded by probe for measuring cervical gene expression.
XX
KW Probe: human; microarray: gene expression; cervical epithelial cell;
XX
KW cervical cancer.
XX
US Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
XX 30-JAN-2001: 2001WO-US00670.
XX
XX 04-FEB-2000: 2000US-0180312.
XX
XX 26-MAY-2000: 2000US-0207456.
XX
XX 30-JUN-2000: 2000US-0608408.
XX
XX 03-AUG-2000: 2000US-0632366.
XX
XX 21-SEP-2000: 2000US-0234687.
XX
XX 27-SEP-2000: 2000US-0236359.
XX
XX 04-OCT-2000: 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SC, Hanzel DK, Chen W, Rank DR;
XX
XX WPI: 2001-488901/53.

Protein involved 1
Human gene 2 encod
Human protein sequ
Equine rhinovirus
Anaplastic carcinoma
C5a peptide analog
Circulin A synthet
Kalata B1 synthet
Human D2H binding
Human polypeptide
Human protein sequ
Human protein sequ
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Corn putative lect
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis Golgi
Arabidopsis thalia
Alfalfa isoflavone
Arabidopsis thalia
Arabidopsis thalia
Human nucleoprotei
Human bone marrow
Human bone marrow
Protein encoded by
Human tumour suppr
Human KCMQ5 (KCM5)
Human KCMQ5 potass
Human myotonic dys
Newcastle disease
S. epidermidis ope
Protein encoded by
Arabidopsis thalia

XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
XX
PS Claim 27: SEQ ID No 21652; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes
CC (SENP: see AAI10068-AA128459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in: grading and/or staging of diseases of the cervix, notably
CC cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 71 AA:

Query Match 90.3%; Score 28; DB 22; Length 71;
Best Local Similarity 66.7%; Pred. No. 77;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPXMR 6
| | | |
Db 1 fhpawr 6

RESULT 2

AA029310
ID AAM29310 standard; Protein: 71 AA.

XX AAM29310;

DT 17-OCT-2001 (first entry)

XX Peptide #3347 encoded by probe for measuring placental gene expression.

DE Probe: microarray: human; placenta; antenatal diagnosis;

KM genetic disorder.

XX Homo sapiens.

OS WO200157272-A2.

XX 09-AUG-2001.

PD 30-JAN-2001; 2001WO-US00663.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488897/53.

XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
XX
PS Claim 27: SEQ ID No 29579; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENP:
CC see AAI13135-AA157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for

CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.

XX Sequence 71 AA:

Query Match 90.3%; Score 28; DB 22; Length 71;
Best Local Similarity 66.7%; Pred. No. 77;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPXMR 6
| | | |
Db 1 fhpawr 6

RESULT 3

AA04538
ID AAM04538 standard; Protein: 71 AA.

XX AAM04538;

DT 09-OCT-2001 (first entry)

DE Peptide #3220 encoded by probe for measuring breast gene expression.

KW Probe: human; breast disease; breast cancer; development disorder;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX Homo sapiens.

OS WO200157270-A2.

XX 09-AUG-2001.

DT 29-JAN-2001; 2001WO-US00661.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-476286/51.

XX Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
XX
XX
PS Claim 27: SEQ ID No 13278; 322pp; English.

XX The present invention relates to novel single exon nucleic acid probes
CC (see AAI10010-AA110067). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for measuring human gene expression in
CC a human breast sample, where the probe hybridises at high stringency to a
CC nucleic acid expressed in the human breast. The probes are useful for
CC predicting, diagnosing, grading, staging, monitoring and prognosing
CC diseases of the human breast, particularly those diseases with polygenic
CC aetiology. The diseases include: breast cancer, disorders of development,
CC inflammatory diseases of the breast, fibrocystic changes, proliferative
CC breast disease and non-carcinoma tumours.

XX Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 71 AA:

Query Match 90.3%; Score 28; DB 22; Length 71;
 Best Local Similarity 66.7%; Pred. No. 77;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPXMR 6
 1 1 1 1
 Db 1 fhpawr 6

RESULT 4

AAB41139
 ID AAB41139, standard; Protein: 277 AA.

AC AAB41139;

DT 08-FEB-2001 (first entry)

XX Human ORFX ORF903 polypeptide sequence SEQ ID NO:1806.

XX Human: open reading frame; ORFX; detection; cytosolic; hepatotropic;
 KM vulnary; antiparkinsonian; nontropic; neuroprotective;
 KM anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KM immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KM hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KM antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;
 KM antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KM neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KM cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KM cholesterol ester storage; systemic lupus erythematosus; infection;
 KM severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KM allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KM bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KM thrombosis; contraceptive.

XX Homo sapiens.

XX MO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000MO-US08621.

XX 31-MAR-1999; 99US-0127607.

XX 02-APR-1999; 99US-0127636.

XX 05-APR-1999; 99US-0127728.

XX 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach M;

XX WPI: 2000-602362/57.

XX N-PSDB: AAC75348.

XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -

XX Claim 11: Page 1407; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytosolic; hepatotropic; vulnary;
 CC antiparkinsonian; nontropic; neuroprotective;
 CC osteopathic; anticonvulsant; antidiabetic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antihypertensive; antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated therapy. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,

CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 277 AA;

Query Match 90.3%; Score 28; DB 21; Length 277;
 Best Local Similarity 66.7%; Pred. No. 2,9e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPXMR 6
 1 1 1 1
 Db 271 fapwyr 276

RESULT 5

AAM24361
 ID AAM24361, standard; Protein: 357 AA.

AC AAM24361;

DT 12-OCT-2001 (first entry)

XX Human EST encoded protein SEQ ID NO: 1886.

XX Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
 KM tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
 KM diagnostics; forensic test; gene mapping; genetic disorder;
 KM biodiversity; gene therapy; nutrition.

XX Homo sapiens.

XX MO200154477-A2.

XX 02-AUG-2001.

XX 25-JAN-2001; 2001MO-US02687.

XX 25-JAN-2000; 2000US-0491404.

XX 17-JUL-2000; 2000US-0617746.

XX 03-AUG-2000; 2000US-0631451.

XX 15-SEP-2000; 2000US-0663870.

XX (HYSE-) HYSPD INC.

XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;

XX Cao Y, Drmanac RA, Zhang J, Werhman T;

XX WPI: 2001-476164/51.

XX N-PSDB: AAM99020.

XX Isolated polypeptide for treatment of diseases, diagnostics, raising
 PT antibodies and research use -

XX Claim 20: Page 1218-1219; 1275pp; English.

XX The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a
 CC protein of the invention.

XX Sequence 357 AA;

Query Match 90.3% Score 28: DB 22: Length 357:
 Best Local Similarity 66.7% Pred. No. 3.8e+02:
 Matches 4: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

OY 1 EXPXMR 6
 1 1 1 1
 241 fspvtr 246

RESULT 6
 AAU00428 :standard: Protein: 383 AA.
 ID AAU00428:
 XX
 AC AAU00428:

14-MAY-2001 (first entry)

Rat Gas1 protein.

Rat: Gas1: lethal cell protein; apoptosis; cell death;
 neuroendocrine disorder; cardiovascular disorder; autoimmune disorder;
 neuroendocrine disorder; oncological disorder; cancer.

Rattus norvegicus.

Key Location/Qualifiers
 Misc-difference 167
 /note="Encoded by TGC"
 Domain 174..279
 /note="Amphipathic alpha-helix region involved in
 channel activity and in death-inducing property
 of Gas1"

WO200114549-A1.

01-MAR-2001.

21-AUG-2000: 2000MO-EP08182.

24-AUG-1999: 99EP-0306702.

(JANC) JANSSEN PHARM NV.

Luyten WHML, Naranjo JR, Wellstroem B:

WPI: 2001-218449/22.

N-PSDB: AAU00428.

Inhibiting lethal effect of protein in cell, involves inhibiting
 function and/or expression of Gas1 protein and expression of sequence
 encoding otherwise lethal protein in the cell

Claim 32: Page 58-59: 65pp: English.

The present sequence represents rat Gas1 protein, a lethal cell protein.
 Novel methods of inhibiting the lethal effect of Gas1 or a functional
 equivalent, derivative or bioprecursor which is capable of inducing
 apoptosis (AAU00429) involve the use of antisense nucleotide sequences
 (AAU00433-AAU00434) which inhibit the expression of the lethal proteins
 in a cell. A nucleotide encoding Gas1 or a protein capable of inducing
 apoptosis, an antisense molecule, Gas1 polypeptide or fragment of, an
 inhibitor or accelerator of cell death, or a pharmaceutical composition
 comprising any of these are useful as a medicament for the prevention or
 treatment of a disease condition such as neurological disorder
 (Parkinson's disease, Alzheimer's disease, Huntington's disease,
 amyotrophic lateral sclerosis, or a neurological condition caused by
 thrombosis or cerebral trauma), cardiovascular disorder (heart attack),
 autoimmune disorder (multiple sclerosis), neuroendocrine disorder
 (neurosis of the pituitary gland), or oncological disorder (cancer),
 mediated at least in part by expression of Gas1 or its functional
 equivalent, derivative or bioprecursor capable of inducing apoptosis in
 a cell or a protein in the pathway of which Gas1 is a component. An

CC antisense molecule is useful as a probe, as a medicament or in the
 CC preparation of a drug for treating the above mentioned diseases.
 XX

Sequence 383 AA:

Query Match 90.3% Score 28: DB 22: Length 383:
 Best Local Similarity 66.7% Pred. No. 4e+02:
 Matches 4: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

OY 1 EXPXMR 6
 1 1 1 1
 Fb 138 fspvtr 143

RESULT 7
 AAY36300 :standard: Protein: 488 AA.
 ID AAY36300:
 XX
 AC AAY36300:

17-SEP-1999 (first entry)

Human secreted protein encoded by gene 77.

Human: secreted protein; cancer; tumour; developmental abnormality;
 foetal deficiency; blood disorder; immune system disorder; inflammation;
 autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
 schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;
 atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;
 digestive disorder; endocrine disorder; infection; AIDS.

Homo sapiens.

WO931117-A1.

24-JUN-1999.

17-DEC-1998: 98MO-US27059.

19-DEC-1997: 97US-0068369.

18-DEC-1997: 97US-0068006.

18-DEC-1997: 97US-0068007.

18-DEC-1997: 97US-0068008.

18-DEC-1997: 97US-0068053.

18-DEC-1997: 97US-0068054.

18-DEC-1997: 97US-0068057.

18-DEC-1997: 97US-0068064.

18-DEC-1997: 97US-0070923.

19-DEC-1997: 97US-0068169.

19-DEC-1997: 97US-0068365.

19-DEC-1997: 97US-0068367.

19-DEC-1997: 97US-0068368.

(HUMA-) HUMAN GENOME SCI INC.

Cartier KC, Duan RD, Feng P, Ferrie AM, Florence C;

Florence K, Greene JM, Janat F, Kyaw H, Moore PA;

NI J, Rosen CA, Ruben SM, Shl Y, Soppet DR, Wei Y;

Ku G;

WPI: 1999-418749/35.

N-PSDB: AAY36300.

New isolated human genes encoding secreted polypeptides

Claim 11: Page 387-389: 537pp: English.

AA97916 to AAY98029 represent 110 isolated human secreted protein

genes. AAY6224 to AAY36727 represent the secreted proteins encoded by

the 110 human genes. The genes and their corresponding secreted

polypeptides are useful for preventing, treating or ameliorating medical

conditions, e.g. by protein or gene therapy. Also pathological conditions

CC can be diagnosed by determining the amount of the new polypeptides in a
 CC sample or by determining the presence of mutations in the new genes.
 CC Specific uses are described for each of the 110 genes, based on which
 CC tissues they are most highly expressed in, and include developing
 CC products for the diagnosis or treatment of cancer, tumours, developmental
 CC abnormalities and foetal deficiencies, blood disorders, diseases of the
 CC immune system, autoimmune diseases, inflammation, allergies, Alzheimer's
 CC and cognitive disorders, schizophrenia, arthritis, asthma, psoriasis,
 CC sepsis, skin disorders, atherosclerosis, diabetes, cardiovascular
 CC disorders, kidney disorders, digestive/endocrine disorders, infections
 CC and AIDS. The polypeptides are also useful for identifying their binding
 CC partners. The sequences given in AAX97907 to AAX97915 and AAY36223 are
 CC used in the exemplification of the present invention.

SQ Sequence 488 AA:

Query Match 90.3%; Score 28; DB 20; Length 488;

Best Local Similarity 66.7%; Pred. No. 5.1e+02;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Caps 0;

OY 1 EXPXMR 6
 Db 241 fsp1wr 246

RESULT 8

AAC92932 AAC92932 standard; Protein: 509 AA.

AC AAG92932:

DT 26-SEP-2001 (first entry)

DE C glutamicum protein fragment SEQ ID NO: 6686.

XX Coryneform bacterium; amino acid synthesis: vitamin; saccharide;

KM organic acid synthesis.

XX Corynebacterium glutamicum.

OS EPI108790-A2.

PN 20-JUN-2001.

PD 18-DEC-2000; 2000EP-0127688.

PF 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX (KYOWA) KYOWA HAKKO KOCYO KK.

PA Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

DR WPI: 2001-376931/40.

XX N-PSDB; AAH68151.

PT Novel polynucleotides derived from Coryneform bacteria, for identifying

PT mutation point of a gene, measuring expression of a gene, analysing

PT expression profile or pattern of a gene and identifying homologous gene

XX Claim 17; SEQ ID NO: 6686; 246pp + Sequence Listing; English.

CC The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Coryneform bacterium, and identifying a homologue of a gene derived
 CC from coryneform bacterium. Coryneform bacteria are useful for producing

CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.

SQ Sequence 509 AA:

Query Match 90.3%; Score 28; DB 22; Length 509;

Best Local Similarity 66.7%; Pred. No. 5.3e+02;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Caps 0;

OY 1 EXPXMR 5
 Db 364 fap1wr 369

RESULT 9

AAB79335 AAB79335 standard; Protein: 509 AA.

AC AAB79335:

DT 30-APR-2001 (first entry)

DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:186.

XX Corynebacterium glutamicum; carbon metabolism and energy production;

KM SMP protein; sugar metabolism and oxidative phosphorylation protein;

KM fine chemical production; organic acid; proteinogenic amino acid;

KM nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;

KM nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;

KM carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;

KM diagnosis; Corynebacterium diptheriae; evolutionary study.

XX Corynebacterium glutamicum.

OS WO200100844-A2.

PN 04-JAN-2001.

PD 23-JUN-2000; 2000WO-1B00943.

PF 25-JUN-1999; 99US-0141031.

PR 08-JUL-1999; 99DE-1031412.

PR 08-JUL-1999; 99DE-1031413.

PR 08-JUL-1999; 99DE-1031419.

PR 08-JUL-1999; 99DE-1031420.

PR 08-JUL-1999; 99DE-1031424.

PR 08-JUL-1999; 99DE-1031428.

PR 08-JUL-1999; 99DE-1031431.

PR 08-JUL-1999; 99DE-1031433.

PR 08-JUL-1999; 99DE-1031434.

PR 08-JUL-1999; 99DE-1031510.

PR 08-JUL-1999; 99DE-1031562.

PR 08-JUL-1999; 99DE-1031634.

PR 09-JUL-1999; 99DE-1032180.

PR 09-JUL-1999; 99DE-1032227.

PR 09-JUL-1999; 99DE-1032308.

PR 09-JUL-1999; 99US-0143208.

PR 14-JUL-1999; 99DE-1032924.

PR 14-JUL-1999; 99DE-1032973.

PR 14-JUL-1999; 99DE-1033005.

PR 27-AUG-1999; 99DE-1040765.

PR 31-AUG-1999; 99US-0151572.

PR 03-SEP-1999; 99DE-1042076.

PR 03-SEP-1999; 99DE-1042079.

PR 03-SEP-1999; 99DE-1042086.

PR 03-SEP-1999; 99DE-1042087.

PR 03-SEP-1999; 99DE-1042088.

PR 03-SEP-1999; 99DE-1042095.

```

PR 03-SEP-1999: 99DE-1042123.
PR 03-SEP-1999: 99DE-1042125.
PA (BAD ) BASF AG.
PI Pompejus M., Kroegeer B., Schroeder H., Zelder O., Haberhauer G.;
XX WPI: 2001-061975/07.
XX N-PSDB: AAF71452.
DR
XX
XX New Isolated Corynebacterium glutamicum nucleic acid encoding a sugar
PT metabolism and oxidative phosphorylation protein for production or
PT modulation of production of fine chemicals e.g. amino acids,
PT carbohydrates or enzymes -
XX
XX Claim 20: Page 407-409; 1246pp; English.
XX
XX AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
CC metabolism and oxidative phosphorylation (SMP) proteins given in
CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
CC energy production. The C. glutamicum SMP gene can be used in vectors
CC (11) for expression in host cells and production or modulation of
CC production of fine chemicals, such as, an organic acid, a proteinogenic
CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
CC cofactor, a polypeptide, or an enzyme. The presence of (I) or SMP proteins
CC (111) encoded by them are used for diagnosing the presence or activity of
CC Corynebacterium diphtheriae in a subject. (I), (11), (111) or host cells
CC containing them are used to map genomes of organisms related to
CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
CC in evolutionary studies, in determining SMP protein regions required
CC for function, in modulating SMP protein activity, in modulating the
CC metabolism of sugars, and in modulating high-energy molecule production
CC in a cell (i.e. ATP, NADPH).
XX
XX Sequence 509 AA:
SO

```

```

Query Match          90.3%: Score 28; DB 22; Length 509;
Best Local Similarity 66.7%: Pred. No. 5.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 1 FXPXMR 6
   | | |
DB 364 fapxmr 369

```

```

RESULT 10
AAW57776
ID AAW57776 standard; Protein: 527 AA.
XX
XX AAW57776:
AC
XX
XX 21-DEC-1998 (first entry)
DT
XX
XX Mycobacterium tuberculosis cellular uptake Mcep protein.
DE
XX
XX Mycobacterium cell entry; Mcep: cellular uptake; tuberculosis;
KM infection; vaccine; therapy; diagnosis; antibody; macrophage.
XX
XX Mycobacterium tuberculosis strain H37Ra (ATCC 25177).
OS
XX
XX Key Location/Qualifiers
FH MISC-difference 66
FT /note= "encoded by AT"
FT MISC-difference 202
FT /note= "encoded by CG"
XX
XX WO9840098-A1.
XX
XX 17-SEP-1998.
XX

```

```

PF 09-MAR-1998: 98MO-US04653.
XX
XX 06-AUG-1997: 97US-0907229.
PR 10-MAR-1997: 97US-0040097.
XX
XX (CORR ) CORNELL RES FOUND INC.
PA
XX
XX Riley LW;
XX
XX WPI: 1998-506483/43.
XX N-PSDB: AAV52613.
DR
XX
XX New Isolated protein from Mycobacterium tuberculosis - useful for
PT e.g. diagnosis, treatment or prevention of infection and for
PT mediating uptake of materials by mammalian cells
XX
XX Claim 9: Page 12-13; 62pp; English.
XX
XX This is the deduced amino acid sequence of a protein encoded by a
CC DNA molecule (see AAV52613) associated with conferring on
CC Mycobacterium tuberculosis an ability to enter mammalian cells.
CC The protein, termed Mcep, or its fragments, especially polypeptides
CC comprising amino acids 1-167, 95-167, 95-528 or 379-528 of the
CC protein, can be expressed in host cells for recombinant production.
CC The proteins are useful in vaccines to prevent infection by M.
CC tuberculosis, and antibodies raised against the proteins can be
CC used for passive immunisation. The proteins, antibodies and DNA
CC molecules may be utilised in diagnostic assays to detect M.
CC tuberculosis in tissue or bodily fluids. The protein can be
CC associated with various other therapeutic materials (e.g.
CC antibiotics, DNA fragments or anti-neoplastic agents) for
CC administration to mammals, particularly humans, to achieve cellular
CC uptake of those materials. The ability of Mcep to augment uptake
CC into macrophages provides an opportunity to deliver genes
CC specifically to macrophages to induce humoral and cell-mediated
CC immunity.
XX
XX Sequence 527 AA:
SO

```

```

Query Match          90.3%: Score 28; DB 19; Length 527;
Best Local Similarity 66.7%: Pred. No. 5.5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 1 FXPXMR 6
   | | |
DB 3 fgpwr 8

```

```

RESULT 11
AAW48359
ID AAW48359 standard; Protein: 527 AA.
XX
XX AAW48359:
AC
XX
XX 03-JUL-1998 (first entry)
DT
XX
XX Mycobacterium bovis BCG protein.
DE
XX
XX Mycobacterium bovis BCG protein; Mcep: tuberculosis; BCGIN-33;
KM hybridisation test; amplification; probe.
XX
XX Mycobacterium bovis.
OS
XX
XX WO9801559-A1.
XX
XX 15-JAN-1998.
XX
XX 09-JUL-1997: 97MO-CA00484.
XX
XX 10-JUL-1996: 96US-0677970.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX

```



```

XX Anand NN, Klein MH;
XX WPI: 1998-110232/10.
DR N-PSDB: AAV17758.
XX
PT Nucleic acid encoding mycobacterial protein involved in cell binding
PT and entry - used for diagnosis of Mycobacterium infection and in
PT vaccines for humans or animals
XX
PS Claim 8; Fig 7A-7E: 107pp: English.
XX
CC The present sequence represents a Mycobacterium bovis BCG protein
CC (45-60 kDa) associated with cell binding and entry. The BCG protein is
CC encoded by the BCGINV-33 gene whose sequence contains two translational
CC frames, one for the M.bovis BCGINV-33 gene and the other for the
CC M.tuberculosis Mce gene. Therefore the Mcep protein is an internal
CC polypeptide of the BCG protein from M.bovis. The invention relates to
CC the use of the BCGINV-33 gene sequence in hybridisation tests for
CC diagnosis of Mycobacterial infections, especially M.tuberculosis
CC infections. Fragments of the BCGINV-33 gene sequence (AAV17758) can be
CC used in PCR to detect Mycobacterium in tissues and body fluids and also
CC for isolating related genes. BCG protein or its fragments can be used
CC in vaccines to generate an immune response that may be capable of
CC protecting humans and animals (especially cattle) against mycobacterial
CC infections.
XX
S0 Sequence 527 AA;

```

Query Match 90.3%; Score 28; DB 19; Length 527;
 Best Local Similarity 66.7%; Pred. No. 5.5e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

OY 1 FXPXMR 6
   1 1 1 1 1
DB 3 fgsprvr 8

```

RESULT 12
 AAY35425
 ID AAY35425 standard; Protein: 629 AA.
 XX
 AC AAY35425;
 XX
 DT 13-SEP-1999 (first entry)
 XX
 DE Protein involved in transcription, translation and/or maturation.
 XX
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW vaccine; neutralising epitope.
 XX
 OS Chlamydia pneumoniae.
 XX
 PN MO9927105-A2.
 XX
 PD 03-JUN-1999.
 XX
 PF 20-NOV-1998; 98WO-1B01890.
 XX
 PR 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97FR-0014673.
 XX
 PA (GEST) GENSET.
 XX
 PI Grlffals R;
 XX
 DR WPI: 1999-357842/30.
 XX
 PT Genome sequence of Chlamydia pneumoniae
 XX
 PS Page 1207-1209; Disclosure: 1912pp; English.

```

XX AAY34584-Y35879 represent the proteins encoded by all the open reading
CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
CC C. pneumoniae causes respiratory disease such as pneumonia and
CC bronchitis and is thought to be a contributing factor in heart
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC nodosum or pharyngitis. The polypeptides encoded by the open reading
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC nucleotide sequences can also be used as immunogenic compositions,
CC especially where the vector directs the expression of a neutralising
CC epitope of C. pneumoniae.
XX
S0 Sequence 629 AA;

```

Query Match 90.3%; Score 28; DB 20; Length 629;
 Best Local Similarity 66.7%; Pred. No. 6.6e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

OY 1 FXPXMR 6
   1 1 1 1 1
DB 290 faplrvr 295

```

RESULT 13
 AAE03765
 ID AAE03765 standard; Protein: 766 AA.
 XX
 AC AAE03765;
 XX
 DT 07-AUG-2001 (first entry)
 XX
 DE Human gene 2 encoded secreted protein HCEJ363, SEQ ID NO:35.
 XX
 KW Human; secreted protein; proliferative disorder; cancer; tumour;
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
 KW cardiovascular disorder; angioinetic disorder; kidney disorder;
 KW gastrointestinal disorder; pregnancy-related disorder; gene therapy;
 KW endocrine disorder; infection; wound healing; vulneryary;
 KW cell culture; chemotaxis; food additive;
 KW binding partner identification.
 XX
 KW Homo sapiens.
 XX
 OS
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..33
 FT Protein /label= signal_peptide
 FT 34..766
 FT /note= "Mature secreted protein"
 XX
 IN MO200132837-A1.
 XX
 PD 10-MAY-2001.
 XX
 PF 17-OCT-2000; 2000WO-US28664.
 XX
 PR 02-NOV-1999; 99US-0163085.
 PR 17-DEC-1999; 99US-0172411.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Piscella M, Wei P, Lafleur DW, Olsen HS, Baker K, Ebner R;
 PI Komatsoulis G, Rosen CA, Ruben SM, Duan RD, Young PE, Florence KA;
 PI Moore PA, Birse CE, Ni J, Soppet DR, Shi Y;
 XX
 DR WPI: 2001-328782/34.
 DR
 NR N-PSDB: AAD08192.
 XX

PT Novel human secreted proteins and nucleic acids for diagnosing,
PT preventing and treating neurological, cardiovascular, infectious,
PT autoimmune, gastrointestinal, bone disorders, cancer, particularly
PT ovarian cancer
PS Claim 11; Page 387-390; 421pp; English.
XX
CC AA008191-AA008213 represent cDNAs corresponding to 13 human secreted
CC protein genes and AA003764-AA003786 represent the proteins they encode.
CC AA003787-AA003800 represent human secreted protein fragments or variants.
CC The genes and their secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy; pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 19 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays (e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA)). The present sequence represents a human
CC secreted protein of the invention.
XX
SQ Sequence 766 AA:

Query Match 90.3%; Score 28; DB 22; Length 766;
Best Local Similarity 66.7%; Pred. No. 8e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EXPXMR 6
1 1 1 1 1
Db 525 fdpsvr 530

RESULT 14
AAB95425
ID AAB95425 standard; Protein: 949 AA.
XX
AC AAB95425;
XX
DT 26-JUN-2001 (first entry)
DE
XX Human protein sequence SEQ ID NO:17833.
KW Human; primer: detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI: 2001-318745/34.
XX
PS
PS Claim 8; SEQ ID 17833; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification, where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 949 AA:

Query Match 90.3%; Score 28; DB 22; Length 949;
Best Local Similarity 66.7%; Pred. No. 9.8e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EXPXMR 6
1 1 1 1 1
Db 241 fsp1vr 246

RESULT 15
AAW27126
ID AAW27126 standard; Protein: 2247 AA.
XX
AC AAW27126;
XX
DT 11-FEB-1998 (first entry)
DE
XX Equine rhinovirus 1 (ErhV1) polypeptide amino acid sequence.
XX
KW Equine rhinovirus 1; ErhV1; foot-and-mouth disease virus; vaccine; horse;
KW diagnosis; antigens; polypeptide; enzyme-linked immunosorbent assay;
XX recombinant protein.
XX
OS Equine rhinovirus 1.
XX
XX
XX Key
XX Cleavage-site 207..208
XX Cleavage-site 287..288
XX Cleavage-site 517..518
XX Cleavage-site 743..744
XX Cleavage-site 991..992

FT Cleavage-site 1007..1008
 FT Cleavage-site 1143..1144
 FT Cleavage-site 1158..1159
 FT Cleavage-site 1552..1553
 FT Cleavage-site 1577..1578
 FT Cleavage-site 1782..1783
 FT Peptide 1..207
 FT Peptide /Label= L
 FT Peptide 208..287
 FT Peptide /Label= VP4
 FT Peptide 288..517
 FT Peptide /Label= VP2
 FT Peptide 518..743
 FT Peptide /Label= VP3
 FT Peptide 744..991
 FT Peptide /Label= VP1
 FT Peptide 992..1007
 FT Peptide /Label= 2A
 FT Peptide 1008..1143
 FT Peptide /Label= 2B
 FT Peptide 1144..1458
 FT Peptide /Label= 2C
 FT Peptide 1459..1552
 FT Peptide /Label= 3A
 FT Peptide 1553..1577
 FT Peptide /Label= 3B
 FT Peptide 1578..1782
 FT Peptide /Label= 3C
 FT Peptide 1783..2246
 FT Peptide /Label= 3D
 PN W09722701-A1.
 XX 26-JUN-1997.
 PD 18-DEC-1996: 96WO-AU00815.
 XX 18-DEC-1995: 95AU-0007201.
 PR (UYME) UNIV MELBOURNE.
 PA
 XX
 PI Crabb BS, Feng L, Studdert MJ:
 XX
 DR WPI: 1997-341692/31.
 N-PSDB: AAT85178.
 XX
 PT Genomic sequence of equine rhinovirus 1 - and derived proteins or
 PT virus-like particles, useful in vaccines and as diagnostic agents
 XX
 PS Claim 2: Pages 30-32: 60pp: English.
 XX
 CC The present sequence represents the amino acid sequence of the
 CC polypeptide of equine rhinovirus 1 (ERHV1). The taxonomic status
 CC of ERHV1 is unclear, as physicochemical studies have shown that the
 CC nucleic acid density and base comparison of ERHV1 differs from other
 CC rhinoviruses. To this end, the nucleotide sequence encoding the
 CC polypeptide of ERHV1 was deduced. Analysis of this sequence suggests that
 CC ERHV1 is more closely related to foot-and-mouth disease virus. Individual
 CC ERHV1 proteins can be used to make vaccines to protect horses (and
 CC possibly other animals) against ERHV1. Oligonucleotide primers and probes
 CC can be used for diagnosis of ERHV1 or related viruses, while antigens of
 CC the ERHV1 polypeptide can be used to detect ERHV1-specific antibodies in
 CC the blood, particularly in enzyme-linked immunosorbent assay. They can
 CC differentiate between infected animals and those vaccinated with ERHV1
 CC vaccines (the infected animals will have antibodies reactive with
 CC non-capsid proteins but vaccinated animals will not). Fragments of the
 CC DNA sequence represent individual genes of the virus and can be
 CC expressed in host systems to produce recombinant proteins. Virus like
 CC particles containing the individual ERHV1 proteins can also be used as
 CC vectors for delivering therapeutic or other useful agents, including
 CC vaccinating epitopes from other pathogens or reproductive hormones.
 XX
 SQ Sequence 2247 AA:

Query Match 90.3%: Score 28: DB 18: Length 2247;
 Best Local Similarity 66.7%: Pred No. 2.3e+03;
 Matches 4: Conservative 0: Mismatches 2: Indels 0: Gaps 0;
 Oy 1 FXPXWR 6
 Db 2227 fvpfwr 2232

Search completed: February 27, 2002, 11:41:19
 Job time: 458 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:45:23 ; Search time 145.23 seconds
(without alignments)
3.147 Million cell updates/sec

Title: US-09-446-109a-17
Perfect score: 31
Sequence: 1 FXPXMR 6

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

1: PIR_68: *
2: PIR1: *
3: PIR2: *
4: PIR3: *
5: PIR4: *

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	93.5	354	2	T39599 conserved hypotet
2	29	93.5	477	2	S77373 hypothetical prote
3	29	93.5	563	2	S78224 virulence-associat
4	29	93.5	639	1	WMV070 70K protein - pola
5	28	90.3	135	2	E82578 hypothetical prote
6	28	90.3	233	1	VMWJBV envelope protein E
7	28	90.3	254	2	S10929 transposase - Myco
8	28	90.3	254	2	A49895 transposase - Flav
9	28	90.3	297	2	E84768 hypothetical prote
10	28	90.3	311	2	T32776 hypothetical prote
11	28	90.3	337	2	B84335 hypothetical prote
12	28	90.3	389	2	G83317 conserved hypotet
13	28	90.3	410	2	S18157 globulin 2 - easte
14	28	90.3	475	1	A69149 O-antigen transpor
15	28	90.3	488	2	S18156 globulin 1 - easte
16	28	90.3	517	2	A70793 probable glycerol
17	28	90.3	542	2	G84911 probable anion exc
18	28	90.3	552	2	G70567 probable ilvB2 pro
19	28	90.3	635	2	G86589 hypothetical prote
20	28	90.3	635	2	G72035 regulatory protein
21	28	90.3	639	2	S03547 hypothetical prote
22	28	90.3	703	2	T48019 hypothetical prote
23	28	90.3	1308	2	B75198 DNA helicase relat
24	27	87.1	89	2	D84256 conserved hypotet
25	27	87.1	113	2	G82517 conserved hypotet
26	27	87.1	119	2	T36433 hypothetical prote
27	27	87.1	199	2	T22304 hypothetical prote
28	27	87.1	236	2	D12355 conserved hypotet
29	27	87.1	252	2	C70380 hypothetical prote

30	27	87.1	267	2	A75339 hypothetical prote
31	27	87.1	301	2	T24993 hypothetical prote
32	27	87.1	303	2	T46416 hypothetical prote
33	27	87.1	309	2	S75505 hypothetical prote
34	27	87.1	310	2	F86737 hypothetical prote
35	27	87.1	318	2	G82350 lipid A biosynthes
36	27	87.1	340	2	G73531 cytochrome c famil
37	27	87.1	345	2	T14707 DNA ligase homolog
38	27	87.1	365	2	T15010 hypothetical prote
39	27	87.1	380	2	G64364 formate hydrogenu
40	27	87.1	395	2	D81146 aminotransferase,
41	27	87.1	395	2	E81877 probable aminotran
42	27	87.1	412	2	E65146 hypothetical 47.4
43	27	87.1	422	2	C82666 conserved hypotet
44	27	87.1	449	2	C84618 hypothetical prote
45	27	87.1	456	2	S61170 BCS1 protein precu

ALIGNMENTS

RESULT 1
*39599
conserved hypothetical protein SPBC16G5.07c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 18-Feb-2000
C:Accession: T39599
R:Lynne, M.; Wood, V.; Rajandream, M.A.; Barrell, B.G.; Beck, A.; Reinhardt, R.
submitted to the EMBL Data Library, May 1998
A:Reference number: 221866
A:Accession: T39599
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-354 <LYN>
A:Cross-references: EMBL:AL023554; PIDN:CAI19027.1; GSPDB:GN00067; SPDB:SPBC16G5.07c
A:Experimental source: Strain 972n-; cosmid c16G5
C:Genetics:
A:Gene: SPDB:SPBC16G5.07c
A:Map position: 2
A:Introns: 72/2: 265/2
C:Superfamily: erythrocyte band 7 integral membrane protein

Query Match 93.5%; Score 29; DB 2; Length 354;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXMR 6
DB 40 FTPXMR 45

RESULT 2

S77373
hypothetical protein s111464 - *Synechocystis* sp. (strain PCC 6803)

C:Species: *Synechocystis* sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999

C:Accession: S77373

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O.; K. Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocys*

5.

A:Reference number: S74322; MUID:97061201

A:Accession: S77373

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-477 <KAN>

A:Cross-references: EMBL:DP0906; GB:AB001339; NID:91652492; PIDN:BAI17476.1; PID:J101

C>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Genetics:

A:Start codon: GTG

Query Match 93.5% Score 29; DB 2; Length 477;
 Best Local Similarity 66.7% Pred. No. 1.7e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Caps 0;

OY 1 FXPXMR 6
 DB 388 FSPSRR 393

RESULT 3
 S78224
 virulence-associated protein mKFB - Salmonella typhimurium plasmid

C:Species: Salmonella typhimurium
 C:Date: 29-Jan-1998 #sequence_revision 13-Feb-1998 #text_change 20-Aug-1999
 C:Accession: S78224; S23713; A43996
 R:Norel, F.
 Submitted to the EMBL Data Library, January 1992
 A:Reference number: S78224
 A:Accession: S78224
 A:Molecule type: DNA
 A:Residues: 1-563 <NOR>
 A:Cross-references: EMBL:X57096; NID:946998; PIDN:CAA0380.1; PID:946999
 R:Norel, F.; Plasco, M.R.; Nicoli, J.; Popoff, M.Y.
 Res. Microbiol. 140, 455-457, 1989
 A:Title: Nucleotide sequence of the plasmid-borne virulence gene mKFB from Salmonella ty
 A:Reference number: A43996; MUID:90161559
 A:Accession: S23713
 A:Molecule type: DNA
 A:Residues: 1-513; 'NITLKGKGRITGVNHFKEAEMLF', 514-563 <NON>
 A:Cross-references: EMBL:X57096
 C:Genetics:
 A:Genome: plasmid
 C:Superfamily: virulence-associated protein spvB
 C:Keywords: DNA binding; transcription regulation

Query Match 93.5% Score 29; DB 2; Length 563;
 Best Local Similarity 66.7% Pred. No. 2e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Caps 0;

OY 1 FXPXMR 6
 DB 50 FAPAMR 55

RESULT 4
 WM070
 70K protein - potato leaf roll virus (strain 1)

C:Species: potato leaf roll virus
 A:Note: host Solanum tuberosum (potato)
 C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 21-Jul-2000
 C:Accession: JA0118; S24591
 R:Mayo, M.A.; Rodinson, D.J.; Jolly, C.A.; Hyman, L.
 J. Gen. Virol. 70, 1037-1051, 1989
 A:Title: Nucleotide sequence of potato leafroll luteovirus RNA.
 A:Reference number: JA0119; MUID:89279282
 A:Accession: JA0118
 A:Molecule type: genomic RNA
 A:Residues: 1-639 <MAY>
 A:Cross-references: EMBL:X14600; NID:9222293; PIDN:BA00417.1; PID:9222297
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1989
 C:Comment: The genome is a single-stranded, positive-sense RNA.
 C:Superfamily: potato leaf roll virus 70K protein

Query Match 93.5% Score 29; DB 1; Length 639;
 Best Local Similarity 66.7% Pred. No. 2.2e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Caps 0;

OY 1 FXPXMR 6
 DB 111111

DB 615 FTPSRR 620

RESULT 5
 E82578
 hypothetical protein XP2271 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: E82578
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MUID:2035717
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: E82578
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-135 <STM>
 A:Cross-references: GB:AE004039; GB:AE003849; NID:99107425; PIDN:AAF85070.1; GSPDB:GN
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer
 as-Neto, F.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to Genbank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Praga, J.S.; Franca, S.C.; Franco, M.C.; Fr
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitchima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
 Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, C.
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sava
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
 M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XP2271

Query Match 90.3% Score 28; DB 2; Length 135;
 Best Local Similarity 66.7% Pred. No. 82;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Caps 0;

OY 1 FXPXMR 6
 DB 79 FVPAMR 84

RESULT 6
 VMW0BV
 envelope protein E - Berne virus

C:Species: Berne virus
 C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 16-Jul-1999
 C:Accession: A39989; S15570
 R:Ben Boon, J.A.; Slijder, E.J.; Locker, J.K.; Horzinek, M.C.; Rottler, P.J.M.
 Virology 182, 655-663, 1991
 A:Title: Another triple-spanning envelope protein among intracellularly budding RNA v
 A:Reference number: A39989; MUID:91220715
 A:Accession: A39989
 A:Molecule type: genomic RNA
 A:Residues: 1-233 <DEN>
 A:Cross-references: GB:X52505; NID:958768; PIDN:CAA36747.1; PID:958769
 R:Boon den, J.A.; Slijder, E.J.; Horzinek, M.C.; Rottler, P.J.M.
 submitted to the EMBL Data Library, February 1990
 A:Reference number: S15570
 A:Accession: S15570
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-233 <BOO>
 A:Cross-references: EMBL:X52505; NID:958768; PIDN:CAA36747.1; PID:958769
 C:Superfamily: Berne virus envelope protein E
 C:Keywords: envelope protein

Query Match 90.3%; Score 28; DB 1; Length 233;
 Best Local Similarity 66.7%; Pred. No. 1.4e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 1 1 1 1
 DB 60 FSPXMR 65

RESULT 7
 S10929
 Transposase - Mycobacterium fortuitum insertion sequence IS6100
 C:Species: Mycobacterium fortuitum
 A:Variety: strain FCI
 C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 24-Sep-1999
 C:Accession: S10929
 R:Martin, C.; Tamm, J.; Rautzier, J.; Gomez-Lus, R.; Davies, J.; Gicquel, D.
 Nature 345, 739-743, 1990
 A:Title: Transposition of an antibiotic resistance element in mycobacteria.
 A:Reference number: S10927; MUID:90294910
 A:Accession: S10929
 A:Molecule type: DNA
 A:Residues: 1-254 <KAT>
 A:Cross-references: EMBL:X53635; NID:g44283; PIDN:CAA37685.1; PID:g44285
 A:Experimental source: strain FCI
 C:Genetics:
 A:Mobile element: Insertion sequence IS6100
 C:Superfamily: Salmonella typhimurium conserved hypothetical protein
 C:Keywords: DNA binding

Query Match 90.3%; Score 28; DB 2; Length 254;
 Best Local Similarity 66.7%; Pred. No. 1.5e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 1 1 1 1
 DB 70 FDPXMR 75

RESULT 8
 A49895
 Transposase - Flavobacterium sp. insertion sequence IS6100
 C:Species: Flavobacterium sp.
 C:Date: 11-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 04-Sep-1998
 C:Accession: A49895
 R:Kato, K.; Ohtsuki, K.; Mitsuoka, H.; Yomo, T.; Negoro, S.; Urabe, I.
 J. Bacteriol. 176, 1197-1200, 1994
 A:Title: Insertion sequence IS6100 on plasmid pOAD2, which degrades nylon oligomers.
 A:Reference number: A49895; MUID:94148782
 A:Accession: A49895
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-254 <KAT>
 A:Cross-references: GB:D26094
 A:Experimental source: strain K172371; plasmid pOAD2; insertion sequence IS6100
 A:Note: nucleotide sequence not given
 C:Genetics:
 A:Mobile element: Insertion sequence IS6100
 C:Superfamily: Salmonella typhimurium conserved hypothetical protein
 C:Keywords: DNA binding

Query Match 90.3%; Score 28; DB 2; Length 254;
 Best Local Similarity 66.7%; Pred. No. 1.5e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 1 1 1 1
 DB 70 FDPXMR 75

RESULT 9

E84788
 Hypothetical protein At2g37100 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: E84788
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Renito, M.I.; Town, C.D.; Fujii, C.Y.
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon,
 cuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487
 A:Accession: E84788
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-297 <STO>
 A:Cross-references: GB:AE002093; NID:g4371291; PIDN:AD018149.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g37100
 A:Map position: 2

Query Match 90.3%; Score 28; DB 2; Length 297;
 Best Local Similarity 66.7%; Pred. No. 1.7e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 1 1 1 1
 DB 118 FSPXMR 123

RESULT 10
 T32776
 Hypothetical protein D1069.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 04-Mar-2000
 C:Accession: T32776
 R:Murray, J.; Langston, Y.; Clarke, K.; Morris, M.
 submitted to the EMBL Data Library, December 1997
 A:Description: The sequence of C. elegans cosmid D1069.
 A:Reference number: 221223
 A:Accession: T32776
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-311 <MUR>
 A:Cross-references: EMBL:AF040641; PIDN:AA94946.1; GSPDB:GN00020; CESP:D1069.1
 A:Experimental source: strain Bristol N2; clone D1069
 C:Genetics:
 A:Gene: CESP:D1069.1
 A:Map position:
 A:Introns: 26477
 C:Superfamily: Caenorhabditis elegans hypothetical protein D1069.1

Query Match 90.3%; Score 28; DB 2; Length 311;
 Best Local Similarity 66.7%; Pred. No. 1.8e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 1 1 1 1
 DB 25 FXPXMR 30

RESULT 11
 B84335
 Hypothetical protein Vng1839h [imported] - Halobacterium sp. NRC-1
 C:Species: Halobacterium sp. NRC-1
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: B84335
 R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky,
 J.; Leitch, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
 Jung, K.H.; Alam, M.; Freitas, T.
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A:Authors: Hou, S.; Daniels, C.D.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.;

A:Title: Genome sequence of Halobacterium species NRC-1.
 A:Reference number: AB4160; MUID:20504483
 A:Accession: B84335
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1337 <STO>
 C:Cross-references: GB:AE004437; NID:g10581284; PIDN:AA020046.1; GSPDB:GN00138
 A:Gene: VNC1839H

Query Match 90.3% Score 28; DB 2; Length 337;
 Best Local Similarity 66.7% Pred. No. 1.9e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 | | | |
 Db 79 FDPXMR 84

RESULT 12
 G83317
 conserved hypothetical protein PA2630 [imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: G83317
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapid, K.; Lim,
 .; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 gen
 A:Reference number: AB2950; MUID:20437337
 A:Accession: G83317
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-389 <STO>
 A:Cross-references: GB:AE004691; GB:AE004091; NID:g9948688; PIDN:AA06018.1; GSPDB:GN001
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA2630

Query Match 90.3% Score 28; DB 2; Length 389;
 Best Local Similarity 66.7% Pred. No. 2.2e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 | | | |
 Db 111 FLPSMR 116

RESULT 13
 S18157
 globulin 2 - eastern white pine
 C:Species: Pinus strobus (eastern white pine)
 C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Aug-1999
 C:Accession: S18157
 R:Rugh, C.L.; Kamalay, J.C.
 submitted to the EMBL Data Library, November 1991
 A:Description: Legumin mRNAs from Pinus strobus L.
 A:Reference number: S18156
 A:Accession: S18157
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-410 <RUG>
 A:Cross-references: EMBL:Z11487; NID:g20744; PIDN:CAA77569.1; PID:g20745
 C:Superfamily: glycinin

Query Match 90.3% Score 28; DB 2; Length 410;
 Best Local Similarity 66.7% Pred. No. 2.3e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 | | | |
 Db 269 FVPSMR 274

RESULT 14
 A69149
 O-antigen transporter related protein - Methanobacterium thermoautotrophicum (strain
 C:Species: Methanobacterium thermoautotrophicum
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
 C:Accession: A69149
 R:Smith, D.R.; Doncelte-Stamm, L.A.; Delonghery, C.; Lee, H.; Dubois, J.; Aldredge, T
 .; Liu, D.; Spadefora, R.; Vitale, R.; Vlatke, Y.; Mierzbowski, J.; Gibson, R.; Jiwani,
 K.L.S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Moelling, J.; Reeve, J.N.
 J. Bacteriol. 179, 7135-7155, 1997
 A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
 A:Reference number: A69000; MUID:98037514
 A:Accession: A69149
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-475 <MTI>
 A:Cross-references: GB:AE000823; GB:AE000666; NID:g2621432; PIDN:AA084885.1; PID:g262
 A:Experimental source: strain Delta H
 C:Genetics:
 A:Gene: MTH379
 A:Start codon: GTG
 C:Superfamily: succinoglycan biosynthesis transport protein

Query Match 90.3% Score 28; DB 1; Length 475;
 Best Local Similarity 66.7% Pred. No. 2.6e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 | | | |
 Db 408 FCPXMR 413

RESULT 15
 S18156
 globulin 1 - eastern white pine
 C:Species: Pinus strobus (eastern white pine)
 C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Aug-1999
 C:Accession: S18156
 R:Rugh, C.L.; Kamalay, J.C.
 submitted to the EMBL Data Library, November 1991
 A:Description: Legumin mRNAs from Pinus strobus L.
 A:Reference number: S18156
 A:Accession: S18156
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-488 <RUG>
 A:Cross-references: EMBL:Z11486; NID:g20742; PIDN:CAA77568.1; PID:g20743
 C:Superfamily: glycinin

Query Match 90.3% Score 28; DB 2; Length 488;
 Best Local Similarity 66.7% Pred. No. 2.7e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 | | | |
 Db 347 FVPSMR 352

Search completed: February 27, 2002, 11:45:23
 Job time: 702 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:42:48 : Search time 78.39 seconds
(without alignments)
2.806 Million cell updates/sec

Title: US-09-446-109A-17

Perfect score: 31
Sequence: 1 FXPXWR 6

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	93.5	452	1 SP52_MOUSE	P97364 mus musculus
2	29	93.5	477	1 YZ64_SYNY3	P73136 synechocyst
3	29	93.5	639	1 V70K_PLRV1	P17519 potato leaf
4	28	90.3	151	1 HSPD_BRAJA	O69241 bradyrhizob
5	28	90.3	151	1 HSPB_BRAJA	O86110 bradyrhizob
6	28	90.3	153	1 HSPB_BRAJA	P70918 bradyrhizob
7	28	90.3	233	1 VENV_BFV	P27904 berne virus
8	28	90.3	333	1 DPSD_CAEEL	Q10949 caenorhabd
9	28	90.3	508	1 GLPK_MYCLE	O9c81 mycobacteri
10	28	90.3	517	1 GLPK_MYCTU	O69654 mycobacteri
11	28	90.3	639	1 V70K_PLRV1	P11622 potato leaf
12	27	87.1	199	1 YAF7_CAEEL	P52883 caenorhabd
13	27	87.1	277	1 HO_FUGHU	O73688 tuqu rubrip
14	27	87.1	456	1 BCS1_YEAST	P32839 saccharomyc
15	27	87.1	459	1 RBL2_RHOSH	P29278 rhodobacter
16	27	87.1	514	1 GSH1_BRAJU	O23736 brassica ju
17	27	87.1	522	1 GSH1_BRAJU	P46309 arabidopsis
18	27	87.1	535	1 YHIL_ECOLI	P37829 escherichia
19	27	87.1	591	1 IF37_ARATH	P56820 arabidopsis
20	27	87.1	623	1 PTR2_CANAL	P46030 candida alb
21	27	87.1	644	1 VP4_BTV2A	P33427 bluetongue
22	27	87.1	878	1 C105_MOUSE	O9145 mus musculu
23	27	87.1	897	1 C105_HUMAN	O9nc82 homo sapien
24	27	87.1	1023	1 TSCC_PSEAM	P55019 pseudopleur
25	27	87.1	2204	1 RRP1_NDVB	P11105 newcastle d
26	26	83.9	224	1 RPIA_LACLA	O9c417 lactococcus
27	26	83.9	407	1 YK67_MYCTU	O10678 mycobacteri
28	26	83.9	486	1 L1P1_YARLI	O99156 yarowia li
29	26	83.9	575	1 TUCA_ECOLI	O47316 escherichia
30	26	83.9	700	1 YHER_ECOLI	P45537 escherichia
31	26	83.9	746	1 TAGP_BACSU	P13485 bacillus su
32	26	83.9	855	1 YB29_YEAST	P38297 saccharomyc
33	26	83.9	3206	1 POLG_PSBMV	P29152 p genome po

34	25	80.6	284	1 YAF1_SYNP1	O05380 synechococc
35	25	80.6	307	1 CRTB_SYNP7	P37269 synechococc
36	25	80.6	493	1 YEB6_YEAST	P39997 saccharomyc
37	25	80.6	497	1 CPD6_HUMAN	P10635 homo sapien
38	25	80.6	497	1 CPDH_MACFA	O29488 macaca fasc
39	25	80.6	497	1 CPDH_CALJA	O18992 callithrix
40	25	80.6	500	1 CPD4_RAT	P13108 ratius norv
41	25	80.6	500	1 CPDG_CAVPO	O64403 cavia porce
42	25	80.6	500	1 CPDI_RAT	O64680 ratius norv
43	25	80.6	523	1 NU4M_PROMI	O37617 prototheca
44	25	80.6	628	1 Y4XM_RHISN	P55706 thizobium s
45	25	80.6	743	1 ANAG_HUMAN	P34802 homo sapien

ALIGNMENTS

RESULT	1	SP52_MOUSE	STANDARD:	PRT:	452 AA.
ID	SP52_MOUSE	P97364:			
AC	15-JUL-1999 (Rel. 38, Created)				
DT	15-JUL-1999 (Rel. 38, Last sequence update)				
DT	30-MAY-2000 (Rel. 39, Last annotation update)				
DE	SELENIDE, WATER DIKINASE 2 (EC 2.7.9.3) (SELENOPHOSPHATE SYNTHETASE 2)				
DE	(SELENIUM DONOR PROTEIN 2).				
CN	SP52.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
OX	NCBI_TaxID=10090;				
RN	11				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=96017645; PubMed=7588067;				
RA	Guinares M.J., Bazan J.F., Zlotnik A., Wiles M.V., Grimaldi J.C.,				
RA	Lee F., McClanahan T.;				
KT	"A new approach to the study of haematopoietic development in the				
KT	yolk sac and embryoid bodies.";				
RL	Development 121:335-346(1995).				
RN	12				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=97140286; PubMed=8986768;				
RA	Guinares M.J., Peterson D., Vicari A., Cocks B.G., Copeland N.G.,				
RA	Gilbert D.J., Jenkins N.A., Ferrick D.A., Kastelein R., Bazan J.F.,				
RA	Zlotnik A.;				
RT	"Identification of a novel selen homology from eukaryotes, bacteria,				
RT	and archaea: is there an autoregulatory mechanism in selenocysteine				
RT	metabolism?";				
RL	Proc. Natl. Acad. Sci. U.S.A. 93:15086-15091(1996).				
CC	- FUNCTION: SYNTHESIZES SELENOPHOSPHATE FROM SELENIUM AND ATP.				
CC	- CATALYTIC ACTIVITY: ATP + SELENIUM + H(2)O = AMP + SELENOPHOSPHATE				
CC	+ PHOSPHATE.				
CC	- COFACTOR: SELENOCYSTEINE. THE ACTIVE-SITE SELENOCYSTEINE IS				
CC	ENCODED BY THE OPAL CODON, UGA.				
CC	- SIMILARITY: BELONGS TO THE SELENOPHOSPHATE SYNTHETASE 1 FAMILY.				
CC	CLASS 1 SUBFAMILY.				
CC	*****				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	*****				
DR	EMBL: U43285; AAC53024.1;				
DR	MED: M01108388; SP52.				
DR	InterPro: IPR000728; AIRS_related.				
DR	Pfam: PF00586; AIRS.1.				
KW	Transferase; Selenium; Selenocysteine; ATP-binding.				
FT	ACT_SITE 63 63				
FT	POTENTIAL 63 63				
FT	SECS 66 66				
FT	*****				
FT	IMPORTANT FOR CATALYTIC ACTIVITY (BY				

```

FT  NP_BIND      322      328      SIMILARITY:
FT  DOMAIN       2        9
FT  DOMAIN      433      440      POLY-ALA.
SO  SEQUENCE     452 AA: 47786 MW: 9DA6F7250CFE80E4 CRC64:

Query Match
Best Local Similarity 66.7%; Score 29; DB 1; Length 452;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY  1 EXPXMR 6
Db  48 FSPSWR 53

RESULT 2
Y264_SYNY3
ID Y264_SYNY3 STANDARD: PRT: 477 AA.
AC P73436:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 54.0 KDA PROTEIN SLL1464.
GN SLL1464.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria: Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneke T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugitara M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Natsu K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT *Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
-1- SIMILARITY: BELONGS TO THE UPF0061 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL: D90906; BAA17476.1;
CC InterPro: IPR003846; UPF0061.
CC Pfam: PF02696; UPF0061.
CC Hypothetical protein: Complete proteome.
SO SEQUENCE 477 AA: 54041 MW: 81F6899B1A6D613C CRC64:

Query Match
Best Local Similarity 93.5%; Score 29; DB 1; Length 477;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPXMR 6
Db 388 FSPSWR 393

RESULT 3
V70K_PLRV1
ID V70K_PLRV1 STANDARD: PRT: 639 AA.
AC P17519:
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)

```

```

DE 69.7 KDA PROTEIN (ORF 2).
OS Potato leafroll virus (strain 1) (PLRV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
OC Polerovirus.
OX NCBI_TaxID=12046;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89279282; PubMed=2732710;
RA Mayo M.A., Robinson D.J., Jolly C.A., Hyman L.;
RT "Nucleotide sequence of potato leafroll luteovirus RNA.";
RL J. Gen. Virol. 70:1037-1051(1989).
-1- SIMILARITY: 54% SIMILARITY TO BEET WESTERN YELLOWS VIRUS ORF2.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL: D00530; BAA00417.1;
CC EMBL: X14600; CAA32741.1;
CC PIR: JAO118; MW070.
CC PIR: S24591; S24591.
CC InterPro: IPR000382; Luteo_ORF2.
CC Pfam: PF02122; Luteo_ORF2.
CC PRINTS: PR00913; LVIRUSORF2.
SO SEQUENCE 639 AA: 69629 MW: AFCF2FB393BEE097 CRC64:

Query Match
Best Local Similarity 93.5%; Score 29; DB 1; Length 639;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPXMR 6
Db 615 FSPSWR 620

RESULT 4
HSPD_BRAJA
ID HSPD_BRAJA STANDARD: PRT: 151 AA.
AC 069241;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SMALL HEAT SHOCK PROTEIN HSPD.
GN HSPD.
OS Bradyrhizobium japonicum.
OC Bacteria: Proteobacteria; alpha subdivision; Rhizobiaceae group;
CC Bradyrhizobium group; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RA Narberhaus F., Weiglhofer W., Fischer H.M., Hennecke H.;
RT "Identification of the Bradyrhizobium japonicum degg gene as part of
RT an operon containing small heat shock protein genes.";
RL Submitted (NOV-1997) to the EMBL/Genbank/DDBJ databases.
-1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL: AJ003064; CAA05835.1;
CC InterPro: IPR002068; Crystallin_HSP20.
CC Pfam: PF00011; HSP20; 1.

```

DR PROSITE: PS01031; HSP20; 1.
KW Heat shock; Multigene family.
SQ SEQUENCE 151 AA; 17272 MW; FB44EF94FB599EE4 CRC64;

Query Match 90.3%; Score 28; DB 1; Length 151;
Best Local Similarity 66.7%; Pred. No. 45;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPXMR 6
| | | | |
DB 6 FSPLMR 11

RESULT 5
HSPH_BRAJA STANDARD; PRT: 151 AA.
AC 086110;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SMALL HEAT SHOCK PROTEIN HSPH.

CN HSPH.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bradyrhizobium group; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.

KA Menchenbach M., Nocker A., Narberhaus F.;
RT Occurrence of a superfamily of small heat shock proteins in
RT Bradyrhizobium japonicum and other Rhizobium species: a plant-like
RT phenomenon.*;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AJ010144; CAA09014.1; -
DR InterPro: IPR002068; Crystallin_HSP20.
DR Pfam: PF00011; HSP20; 1.

DR PROSITE: PS01031; HSP20; 1.
KW Heat shock; Multigene family.
SQ SEQUENCE 151 AA; 17098 MW; DFE98093110AF6A28 CRC64;

Query Match 90.3%; Score 28; DB 1; Length 151;
Best Local Similarity 66.7%; Pred. No. 45;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPXMR 6
| | | | |
DB 6 FSPLMR 11

RESULT 6
HSPB_BRAJA STANDARD; PRT: 153 AA.
AC P70918;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SMALL HEAT SHOCK PROTEIN HSPB.

CN HSPB.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

CC Bradyrhizobium group; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=96404779; PubMed=8808920;

RA Narberhaus F., Weiglhofer W., Fischer H.M., Hennecke H.;
RT "The Bradyrhizobium japonicum rpoH1 gene encoding a sigma 32-like
RT protein is part of a unique heat shock gene cluster together with
RT groESL and three small heat shock genes.*";
RL J. Bacteriol. 178:5337-5346(1996).

CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U55047; AAC44756.1; -
DR InterPro: IPR002068; Crystallin_HSP20.
DR Pfam: PF00011; HSP20; 1.

DR PROSITE: PS01031; HSP20; 1.
KW Heat shock; Multigene family.
SQ SEQUENCE 153 AA; 17170 MW; E2E7E248A69BFD38 CRC64;

Query Match 90.3%; Score 28; DB 1; Length 153;
Best Local Similarity 66.7%; Pred. No. 45;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPXMR 6
| | | | |
DB 7 FAPLMR 12

RESULT 7
VENV_BEV STANDARD; PRT: 233 AA.
ID VENV_BEV
AC P27904;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE ENVELOPE PROTEIN.

DR EMBL: A010144; CAA09014.1; -
DR InterPro: IPR002068; Crystallin_HSP20.
DR Pfam: PF00011; HSP20; 1.

DR PROSITE: PS01031; HSP20; 1.
KW Heat shock; Multigene family.
SQ SEQUENCE 151 AA; 17098 MW; DFE98093110AF6A28 CRC64;

Query Match 90.3%; Score 28; DB 1; Length 153;
Best Local Similarity 66.7%; Pred. No. 45;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPXMR 6
| | | | |
DB 7 FAPLMR 12

RESULT 7
VENV_BEV STANDARD; PRT: 233 AA.
ID VENV_BEV
AC P27904;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE ENVELOPE PROTEIN.

CN E.
OS Herpes virus (BEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Torovirus.
OX NCBI_TaxID=11156;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-ISOLATE: P138/72;
RX MEDLINE=91220715; PubMed=2024492;
RA den Boon J.A., Snijder E.J., Locker J.K., Horzinek M.C.,
FA Rotter P.J.M.;
RT "Another triple-spanning envelope protein among intracellularly
RT budding RNA viruses: the torovirus E protein.*";
RL Virology 182:655-663(1991).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X52505; CAA36747.1; -
DR PIR: A39989; VMMOBY.
DR PIR: A39989; VMMOBY.
DR PIR: A39989; VMMOBY.
KW Envelope protein.

SO SEQUENCE 233 AA: 26548 MW: A07A34DC539104BE CRC64:

Query Match 90.3%; Score 28; DB 1; Length 233;
Best Local Similarity 66.7%; Pred. No. 67;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
1 1 1 1
DB 60 FSPXMR 65

RESULT 8

DPSD_CAEBL STANDARD: PRT: 333 AA.

AC 010949;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE PUTATIVE PHOSPHATIDYL-SERINE DECARBOXYLASE PROENZYME (EC 4.1.1.65).

GN B0361.5.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RA 111

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Du 2.;

RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.

CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYL-L-SERINE =

PHOSPHATIDYLETHANOLAMINE + CO(2).

CC -1- COFACTOR: REQUIRES A PYRUVYL GROUP FOR ITS ACTIVITY

(BY SIMILARITY).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC EMBL: U00031; AKI8866.1; -

DR WormRep: B0361.5; CE00834.

DR InterPro: IPR003817; PS_Dcarboxylase.

DR Pfam: PF02666; PS_Dcarboxylase; 1.

KW Hypothetical protein: Phospholipid biosynthesis; Lyase; Decarboxylase;

KW Pyruvate; Zymogen.

FT CHAIN 1 299

FT CHAIN 300 333

FT SITE 299 300

FT MOD.RES 300 300

FT SEQUENCE 333 AA: 37596 MW: 18CF04FEF31E1F34 CRC64:

Query Match 90.3%; Score 28; DB 1; Length 333;
Best Local Similarity 66.7%; Pred. No. 93;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
1 1 1 1
DB 36 FXPXMR 41

RESULT 9

GLPK_MYCLE STANDARD: PRT: 508 AA.

AC 09CB81;

DT 20-AUG-2001 (Rel. 40, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

DE GLYCEROL KINASE (EC 2.7.1.30) (ATP:GLYCEROL 3-PHOSPHOTRANSFERASE)

DE (GLYCEROKINASE) (CK).

GN GLPK OR ML2314.

OS Mycobacterium leprae.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxID=1769;

RA 111

RP SEQUENCE FROM N.A.

RC STRAIN-TN;

RX MEDLINE=21128732; PubMed=11234002;

RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,

RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,

RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,

RA Davies R.M., Devlin K., Duthoy S., Fellwell T., Fraser A., Hamlin N.,

RA Holtroyd S., Hornsby T., Jagels K., Jancovic C., Maclean J., Moule S.,

RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,

RA Rutter S., Seeger K., Simon S., Simmonds M., Skellern J., Squares R.,

RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,

PA Barrell B.G.;

ET "Massive gene decay in the leprosy bacillus.";

RL Nature 409:1007-1011(2001).

CC -1- FUNCTION: KEY ENZYME IN THE REGULATION OF GLYCEROL UPTAKE AND

METABOLISM.

CC -1- CATALYTIC ACTIVITY: ATP + GLYCEROL = ADP + GLYCEROL 3-PHOSPHATE.

CC -1- PATHWAY: RATE-LIMITING STEP IN GLYCEROL UTILIZATION.

CC -1- SIMILARITY: BELONGS TO THE PUCOKINASE / GLUCOKINASE /

GLYCEROKINASE / XYLULOXINASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC EMBL: AL583925; CAC31830.1; -

DR Leproma; MU2314; -

DR InterPro: IPR000577; FGGY_kin.

DR Pfam: PF00370; FGGY_1.

DR PROSITE: PS00933; FGGY_KINASES_1; 1.

DR PROSITE: PS00445; FGGY_KINASES_2; 1.

KW Glyceral metabolism; Transferrase; Kinase; ATP-binding;

KW Complete proteome.

FT NE_BIND 157 169

FT SEQUENCE 508 AA: 54658 MW: A15379793ECF4039 CRC64:

Query Match 90.3%; Score 28; DB 1; Length 508;
Best Local Similarity 66.7%; Pred. No. 1,4e+02;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
1 1 1 1
DB 360 FSPXMR 365

RESULT 10
GLPK_MYCTU STANDARD: PRT: 517 AA.

AC 069664;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE GLYCEROL KINASE (EC 2.7.1.30) (ATP:GLYCEROL 3-PHOSPHOTRANSFERASE)

DE (GLYCEROKINASE) (CK).

GN GLPK OR RV3696C OR MT3798 OR MTW025.044C.

OS Mycobacterium tuberculosis.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
 Horsby T., Jags K., Krogh A., McLean J., Moule S., Murphy L.,
 Oliver K., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 Ruster S., Seeger K., Skelton J., Squares R., Squares S.,
 Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
 RA "Deciphering the biology of mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 RN 12)
 RP SEQUENCE FROM N.A.
 RC STRAIN-CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 Peterson J., Deboy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
 Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 Bishai W., Ulfersack T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: KEY ENZYME IN THE REGULATION OF GLYCEROL UPTAKE AND
 CC METABOLISM.
 CC -1- CATALYTIC ACTIVITY: ATP + GLYCEROL = ADP + GLYCEROL 3-PHOSPHATE.
 CC -1- PATHWAY: RATE-LIMITING STEP IN GLYCEROL UTILIZATION.
 CC -1- SIMILARITY: BELONGS TO THE FUCOKINASE / GLUCOKINASE /
 CC GLYCEROKINASE / XYLOKINASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC
 CC EMBL: AL022121; CA18018.1; -
 DR EMBL: AE007177; AAK48165.1; -
 DR HSSP: P08859; 1GLB.
 DR TIGR: MT3798; -
 DR Tuberculist: R3696C; -
 DR InterPro: IPR000577; FCGY_kin.
 DR Pfam: PF00370; FCGY_1.
 DR PROSITE: PS00445; FCGY_KINASES_2; 1.
 DR PROSITE: PS00933; FCGY_KINASES_1; 1.
 KM Glycerol metabolism; Transferase; kinase; ATP-binding;
 KW Complete proteome.
 FT NP_BIND 165 177 ATP (PROBABLE).
 SO SEQUENCE 517 AA; 55859 MW; 7E3F30DCFA2E63E CRC64;

UT 01-AUG-1990 (Rel. 15, Last annotation update)
 DE 69.7 KDA PROTEIN (ORF 2).
 OS Potato leafroll virus (strain Wageningen) (PlRV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
 OC Polerovirus.
 CX NCBI_TaxID=12048;
 RN 11)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89171329; PubMed=2466700;
 RA van der Milk F., Huisman M.J., Cornelissen B.J.C., Huttinga H.,
 RA Goldbach R.W.;
 RT "Nucleotide sequence and organization of potato leafroll virus
 RT genomic RNA."
 RL FEBS Lett. 245:51-56(1989).
 CC -1- SIMILARITY: 54% SIMILARITY TO BEET WESTERN YELLOWS VIRUS ORF2.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC
 CC EMBL: Y07496; CA68795.1; -
 DR PIR: S03547; S03547.
 DR InterPro: IPR000382; Luteo_ORF2.
 DR Pfam: PF02122; Luteo_ORF2; 1.
 DR PRINTS: PR00913; LVIRUSORF2.
 SO SEQUENCE 639 AA; 69676 MW; 92E1473FE3FE148 CRC64;

Query Match 90.3%; Score 28; DB 1; Length 639;
 Best Local Similarity 66.7%; Pred. No. 1.7e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EXPXMR 6
 DB 615 FIPSWR 620

RESULT 12
 ID YAF7_CAEEL STANDARD; PRT: 199 AA.
 AC P52883;
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE HYPOTHETICAL 23.2 KDA PROTEIN F46C5.7 IN CHROMOSOME II.
 GN F46C5.7.
 CS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 CC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN 11)
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Thomas K.;
 RL Submitted (SEP-1995) to the EMBL/Genbank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC
 CC EMBL: Z54281; CA91049.1; -
 DR WormRep: F46C5.7; GE03348.
 KW Hypothetical protein; transmembrane.
 FT TRANSMEM 35 55 POTENTIAL.

FT TRANSMEM 57 77 POTENTIAL.
 FT TRANSMEM 94 114 POTENTIAL.
 FT TRANSMEM 131 151 POTENTIAL.
 SQ SEQUENCE 199 AA: 23247 MW: AC09222A692E3638 CRC64:

Query Match 87.1%; Score 27; DB 1; Length 199;
 Best Local Similarity 66.7%; Pred. No. 90;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 EXPXMR 6
 Db 22 EXPXMR:27

RESULT 13
 HO_FUGRU STANDARD; PRT: 277 AA.
 AC 073688;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE HEME OXYGENASE (EC 1.14.99.3) (HO).
 OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontidae; Takifugu.
 RX NCBI_TaxID=31033;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=98163746; PubMed=9503016;
 RA Goltgens B., Gilbert J.G.R., Barton L.M., Aparicio S., Hawker K.,
 RA Mistry S., Vaubin M., King A., Bentley D., Elgar G., Green A.R.;
 RT "The pufferfish SLP-1 gene, a new member of the SCL/TAL-1 family of
 RT transcription factors";
 RL Genomics 48:52-62(1998).

CC -!- FUNCTION: HEME OXYGENASE CLEAVES THE HEME RING AT THE ALPHA
 CC MENENGE BRIDGE TO FORM BILIVERDIN. BILIVERDIN IS SUBSEQUENTLY
 CC CONVERTED TO BILIRUBIN BY BILIVERDIN REDUCTASE. UNDER
 CC PHYSIOLOGICAL CONDITIONS, THE ACTIVITY OF HEME OXYGENASE IS
 CC HIGHEST IN THE SPLEEN, WHERE SENESCENT ERYTHROCYTES ARE
 CC SEQUESTERED AND DESTROYED (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: HEME + 3 AH(2) + O(2) -> BILIVERDIN + FE(2+) +
 CC CO + 3 A + 3 H(2)O.
 CC -!- SUBCELLULAR LOCATION: MICROSOMAL (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE HEME OXYGENASE FAMILY.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC EMBL: AF022814; AAC41263.1; -
 DR InterPro: IPR002051; Heme_oxygenase.
 DR Pfam: PF01126; Heme_oxygenase; 1.
 DR PRINTS: PR00088; HAEMOXYGNASE.
 DR PROSITE: PS00593; HEME_OXYGENASE; 1.
 DR Heme; Oxidoreductase; Microsome.
 KW BINDING 29
 FT SEQUENCE 277 AA: 31211 MW: 7783584699963677 CRC64;

Query Match 87.1%; Score 27; DB 1; Length 277;
 Best Local Similarity 66.7%; Pred. No. 1.2e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 EXPXMR 6
 Db 101 EXPXMR 106

RESULT 14
 BCS1_YEAST STANDARD; PRT: 456 AA.
 AC P32839; 006404;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE BCS1 PROTEIN.
 GN BCS1 OR YDR375C OR D9481.17.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Saccharomycetes; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Eukaryota; Fungi; Ascomycota; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 RX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=93010976; PubMed=1327750;
 RA Nobrega F.G., Nobrega M.P., Tzagoloff A.;
 RT "BCS1, a novel gene required for the expression of functional Rieske
 RT iron-sulfur protein in Saccharomyces cerevisiae.";
 RL EMBO J. 11:3821-3829(1992).

CC -!- FUNCTION: ESSENTIAL FOR THE EXPRESSION OF THE RIESKE IRON-SULFUR
 CC PROTEIN.
 CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND. MITOCHONDRIAL.
 CC -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
 CC CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC EMBL: S47190; AAC09007.1; -
 DR EMBL: U28373; AAB64811.1; -
 DR PIR: S28411; S28411.
 DR SCD: S0002783; BCS1.
 DR InterPro: IPR003593; AAA.
 DR InterPro: IPR001939; AAA_subfam.
 DR Pfam: PF00004; AAA; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00674; AAA; 1.
 KW ATP-binding; Mitochondrion; Transmembrane.
 FT DOMAIN 1 44 MITOCHONDRIAL INTERMEMBRANE (PROBABLE).
 FT TRANSMEM 69 68 PROBABLE.
 FT SITE 69 83 MITOCHONDRIAL MATRIX (PROBABLE).
 FT NP_BIND 267 274 ATP (POTENTIAL).
 FT CONFLICT 119 119 N -> F (IN REF. 1).
 FT CONFLICT 267 267 G -> S (IN REF. 1).
 FT CONFLICT 349 349 L -> Q (IN REF. 1).
 FT CONFLICT 351 351 G -> S (IN REF. 1).
 SQ SEQUENCE 456 AA: 51107 MW: 177500268664ED04 CRC64;

Query Match 87.1% Score 27; DB 1; Length 456;
 Best Local Similarity 66.7% Pred. No. 1.9e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXMR 6
 1 1 1 1
 DB 209 FGPXMR 214

RESULT 15

RL2_RHOSH STANDARD; PRT; 459 AA.
 ID RL2_RHOSH
 AC P29278;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN 2 (EC 4.1.1.39)
 DE (RUBISCO LARGE SUBUNIT).
 GN CBBL2 OR RBPL.
 OS Rhodospirillum rubrum (Rhodospirillum rubrum).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum group;
 OC Rhodospirillum
 NC NCBI_TaxID=1063;
 RN 111
 RP SEQUENCE FROM N.A.
 RA Wagner S.J., Stevens S.E. Jr., Nixon B.T., Lambert D.H.,
 RA Oliver R.G. Jr., Tabita F.R.;
 RT "Nucleotide and deduced amino acid sequence of the Rhodospirillum
 RT sphaeroides gene encoding form II ribulose-1,5-bisphosphate
 RT carboxylase/oxygenase and comparison with other deduced forms I and II
 RT sequences.";
 RL FEMS Microbiol. Lett. 55:217-222(1988).
 RN 121
 RP SEQUENCE OF 1-36 FROM N.A.
 RX MEDLINE=92041881; Pubmed=1939098;
 RA Chen J.-H., Gibson J.L., McCue L.A., Tabita F.R.;
 RT "Identification, expression, and deduced primary structure of
 RT transketolase and other enzymes encoded within the form II CO2
 RT fixation operon of Rhodospirillum rubrum.";
 RL J. Biol. Chem. 266:20447-20452(1991).
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
 CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
 CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
 CC ACTIVE SITE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) =
 CC 2 3-PHOSPHO-D-GLYCERATE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) =
 CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
 CC -1- SUBUNIT: IN CONTRAST TO FORM I GENES, THE FORM II RUBISCO ARE
 CC COMPOSED OF ONLY LARGE SUBUNITS.
 CC -1- MISCELLANEOUS: THIS PROTEIN IS ENCODED WITHIN THE FORM II
 CC RIBULOSE-BISPHOSPHATE CARBOXYLASE OPERON.
 CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M68914; AAA26158.1; ALT_SEQ.
 CC PIR: E41080; E41080.
 CC HSSP: P04718; 1RBA.
 CC InterPro: IPR000685; Rubisco_large.
 CC Pfam: PF00016; Rubisco_large.1.
 CC PROSITE: PS00157; RUBISCO_LARGE; 1.
 CC Photosynthesis; Carbon dioxide fixation; Photorespiration;

KW Lyase; Oxidoreductase; Monooxygenase; Multigene family.
 FT ACT_SITE 191 191 BINDING OF CO(2) ACTIVATES THE ENZYME.
 SQ SEQUENCE 459 AA; 50519 MW; 299ABAA836BD683E CRC64;

Query Match 87.1% Score 27; DB 1; Length 459;
 Best Local Similarity 66.7% Pred. No. 1.9e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXMR 6
 1 1 1 1
 DB 447 FGPXMR 452

Search completed: February 27, 2002, 11:42:48
 Job time: 547 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:50:14 ; Search time 281.76 Seconds
(without alignments)
3.115 Million cell updates/sec

Title: US-09-446-109a-17
Sequence: 1 FXPXMR 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: SP archaea:*
- 2: SP bacteria:*
- 3: SP fungi:*
- 4: SP human:*
- 5: SP invertebrate:*
- 6: SP mammal:*
- 7: SP mhc:*
- 8: SP organelle:*
- 9: SP phage:*
- 10: SP plant:*
- 11: SP rodent:*
- 12: SP virus:*
- 13: SP vertebrate:*
- 14: SP unclassified:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	93.5	296	2	033353 mycobacteri
2	29	93.5	354	3	060121 schizosacch
3	29	93.5	563	2	P72421 salmonella
4	29	93.5	572	12	093886
5	28	90.3	135	2	09PR73
6	28	90.3	141	2	09WWS5
7	28	90.3	143	5	096402
8	28	90.3	160	4	095530
9	28	90.3	160	4	098029
10	28	90.3	211	2	033867
11	28	90.3	218	4	095726
12	28	90.3	221	10	09XJ12
13	28	90.3	236	2	005115
14	28	90.3	241	5	09VLD3
15	28	90.3	250	2	09EUM2
16	28	90.3	254	2	049185
17	28	90.3	254	2	09F1E6
18	28	90.3	264	2	09FDH9
19	28	90.3	272	11	09DA91

20	28	90.3	297	10	09ZOC7	09ZOC7 arabidopsis
21	28	90.3	311	5	044789	044789 caenorhabdl
22	28	90.3	317	2	09L4F0	09L4F0 bacillus ce
23	28	90.3	325	2	09L4E9	09L4E9 bacillus ce
24	28	90.3	337	1	09HP24	09HP24 halobacteri
25	28	90.3	357	5	026862	026862 trypanosoma
26	28	90.3	389	2	0910K8	0910K8 pseudomonas
27	28	90.3	407	2	09AAS9	09AAS9 caulobacter
28	28	90.3	410	10	041018	041018 pinus strob
29	28	90.3	475	1	026479	026479 methanobact
30	28	90.3	483	2	09A8C6	09A8C6 caulobacter
31	28	90.3	488	10	041017	041017 pinus strob
32	28	90.3	490	2	09A3L3	09A3L3 caulobacter
33	28	90.3	507	2	09RJM2	09RJM2 streptomyce
34	28	90.3	512	2	09ADA7	09ADA7 streptomyce
35	28	90.3	542	10	09SHX0	09SHX0 arabidopsis
36	28	90.3	552	2	006335	006335 mycobacteri
37	28	90.3	599	4	09HSP1	09HSP1 homo sapien
38	28	90.3	624	4	09C0D7	09C0D7 homo sapien
39	28	90.3	633	4	09Y4W3	09Y4W3 homo sapien
40	28	90.3	635	2	0927B3	0927B3 chlamydia p
41	28	90.3	635	2	09JSA3	09JSA3 chlamydia p
42	28	90.3	639	12	084836	084836 potato leaf
43	28	90.3	703	10	09M1P7	09M1P7 arabidopsis
44	28	90.3	781	4	095560	095560 homo sapien
45	28	90.3	791	4	09C0B6	09C0B6 homo sapien

ALIGNMENTS

```

RESULT 1
ID 033353 PRELIMINARY: PRT: 296 AA.
AC 033353;
BT 01-JAN-1998 (TREMBLrel, 05, Created)
DT 01-JUN-1998 (TREMBLrel, 06, Last sequence update)
NT 01-MAY-2000 (TREMBLrel, 13, Last annotation update)
DE HYPOTHETICAL 32.0 KDA PROTEIN (ORF2).
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-IS1547;
KC MEDLINE=99121044; PubMed=9922269;
RA Fang Z., Doig C., Morrison N., Watt B., Forbes K.J.;
RT "Characterization of IS1547, a new member of the IS900 family in the
RT Mycobacterium tuberculosis complex, and its association with IS6110.";
RL J. Bacteriol. 181:1021-1024(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TRANSPOSON-IS1547;
KC MEDLINE=97157336; PubMed=9003621;
RA Fang Z., Forbes K.J.;
RT "A Mycobacterium tuberculosis IS6110 preferential locus (lpl) for
RT insertion into the genome.";
RL J. Clin. Microbiol. 35:479-481(1997).
DR EMBL: Y13470; CAA73869.1;
KW Transposable element; Hypothetical protein.
SQ SEQUENCE 296 AA; 31996 MW; 3CDACBC9566C53F0 CRC64;

```

Query Match 93.5%; Score 29; DB 2; Length 296;
Best Local Similarity 66.7%; Pred. No. 3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 FXPXMR 6
263 FSPSWR 268

```

RESULT 2
060121 ID 060121 PRELIMINARY: PRT: 354 AA.
AC 060121:
DT 01-AUG-1998 (TEMBLrel. 07, Created)
DT 01-AUG-1998 (TEMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE HYPOHETICAL 39.3 KDA PROTEIN C16G5.07C IN CHROMOSOME II.
GN SPBC16G5.07C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Lyne M., Wood V., Rajandream M.A., Barrell B.G., Beck A.,
RA Reinhardt R.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE BAND 7 / MEC-2 FAMILY.
DR EMBL: AL023554; CAA19027.1; -.
DR InterPro: IPR001107; Band_7.
DR InterPro: IPR001972; Stomatlin.
DR Pfam: PF01145; Band_7; 1.
DR SMART: SM00244; PHB; 1.
DR PROSITE: PS01270; BAND_7; FALSE_NEG.
DR Hypothetical protein.
KW SEQUENCE 354 AA; 39274 MW; 66B09E6A12BDC030 CRC64;
SQ

Query Match 93.5%; Score 29; DB 3; Length 354;
Best Local Similarity 66.7%; Pred. No. 3.5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

UY 1 FXPXMR 6
| | |
Db 40 FTPTWR 45

RESULT 3
P72421 ID P72421 PRELIMINARY: PRT: 563 AA.
AC P72421:
DT 01-FEB-1997 (TEMBLrel. 02, Created)
DT 01-FEB-1997 (TEMBLrel. 02, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE (SEROTYPE TYPHIMURIUM).
GN MKFB.
OS Salmonella typhimurium.
OC Plasmid.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SEROTYPE TYPHIMURIUM, C5;
RC MEDLINE=90161559; PubMed=2696057;
RA Norel F., Pisano M.R., Nicoll J., Popoff M.Y.;
RT "Nucleotide sequence of the plasmid-borne virulence gene mltB from
RT Salmonella typhimurium";
RL Res. Microbiol. 140:455-457(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SEROTYPE TYPHIMURIUM, C5;
RC Norel F.;
RA Submitted (JAN-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL: X57096; CAA40380.1; -.
DR InterPro: IPR003284; Sal_SpVb.
DR PRINTS: PR01341; SALSPVBPROT.
KW Plasmid.
SQ SEQUENCE 563 AA; 62586 MW; DA3F093AB8241C0A CRC64;

```

```

Query Match 93.5%; Score 29; DB 2; Length 563;
Best Local Similarity 66.7%; Pred. No. 5.5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

UY 1 FXPXMR 6
| | |
Db 50 FAPXMR 55

RESULT 4
Q9J8B6 ID Q9J8B6 PRELIMINARY: PRT: 572 AA.
AC Q9J8B6:
DT 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE OREF19 CHITINASE.
OS Spodoptera exigua nucleopolyhedrovirus.
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
NCBI_TaxID=10454;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20036646; PubMed=10567663;
RA Ikel W.F., van Strien E.A., Heldens J.G., Broer R., Zuidema D.,
RA Goldbach R.W., Vlak J.M.;
RT "Sequence and organization of the spodoptera exigua multicapsid
RT nucleopolyhedrovirus genome.";
PL J. Gen. Virol. 80:3289-3304(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Ikel W.F., van Strien E.A., Heldens J.G.M., Broer R., Zuidema D.,
RA Goldbach R.W., Vlak J.M.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF169823; AAF33549.1; -.
DR InterPro: IPR000886; ER_target.
DR InterPro: IPR001223; Glyco_hydro_18.
DR InterPro: IPR002173; pfb.
DR Pfam: PF00704; Glyco_hydro_18; 1.
DR PROSITE: PS00014; ER_TARGET; UNKNOWN_1.
DR PROSITE: PS00583; PFB_KINASES_1; UNKNOWN_1.
KW SEQUENCE 572 AA; 63508 MW; A678ADB14021A014 CRC64;
SQ

Query Match 93.5%; Score 29; DB 12; Length 572;
Best Local Similarity 66.7%; Pred. No. 5.5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

UY 1 FXPXMR 6
| | |
Db 401 FAPXMR 406

RESULT 5
Q9PB73 ID Q9PB73 PRELIMINARY: PRT: 135 AA.
AC Q9PB73:
DT 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 16, Last annotation update)
DE HYPOHETICAL PROTEIN XF2271.
GN XF2271.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RA MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,

```

RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
 RA Colauco N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Facincal A.P., Ferreira A.J.S., Ferreira V.C.A., Ferto J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 RA Garner M., Goldman G.H., Goldman H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Honeisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Montello-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nham A.Jr., Nobrega F.C., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peloto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Zeldanis J., Zetubal J.C.,
 RT The genome sequence of the plant pathogen *Xylella fastidiosa*.
 RL Nature 406:151-159(2000).
 DR EMBL: AEO04039; AAF85070.1; -
 KM Hypothetical protein: Complete proteome.
 SQ SEQUENCE 135 AA: 15203 MW: 99EF0BA222D5244 CRC64;

Query Match 90.3% Score 28; DB 2; Length 135;
 Best Local Similarity 66.7% Pred. No. 2.2e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXMR 6
 DB 79 FVPMWR 84

RESULT 6
 Q9WMK5 PRELIMINARY: PRT: 141 AA.

AC 09WMK5: 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE HYPOTHETICAL 16.0 KDA PROTEIN.
 OS *Synechococcus* sp. (strain PCC 7002) (Agmenellum quadruplicatum).
 OC Bacteria; Cyanobacteria; Chroococcales; *Synechococcus*.
 OX NCBI_TaxID=32049;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PCC7002; PR6000;
 RA Schlink K., Kowalik K.V., Bryant D.A.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF136577; AAD26588.1; -
 DR HSSP: P00210; 1FXR.
 DR InterPro: IPR01080; 3FeS_ferredox.
 DR PRINTS: PR00352; 3FESFRODXT.
 KW Hypothetical protein.
 SQ SEQUENCE 141 AA: 16034 MW: 85872F258B76FCE2 CRC64;

Query Match 90.3% Score 28; DB 2; Length 141;
 Best Local Similarity 66.7% Pred. No. 2.3e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXMR 6
 DB 122 FSPQWR 127

RESULT 7
 C96402 PRELIMINARY: PRT: 143 AA.

AC 096402: 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
 DT 01-MAY-1999 (TREMblrel. 10, Last annotation update)
 DE ERYTHROCYTE MEMBRANE PROTEIN 1 (FRAGMENT).
 GN FC93S1.2VAR6
 OS *Plasmodium falciparum*.
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Chen Q., Mahlgren M.;
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF039278; AAD02161.1; -
 FT NON-TER 1 143
 FT NON-TER 1 143
 SQ SEQUENCE 143 AA: 16818 MW: A8641F68BE09B3CE CRC64;

Query Match 90.3% Score 28; DB 5; Length 143;
 Best Local Similarity 66.7% Pred. No. 2.4e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXMR 6
 DB 2 FNPTWR 7

RESULT 8
 O95530 PRELIMINARY: PRT: 160 AA.

AC 095530: 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
 DT 01-MAY-1999 (TREMblrel. 10, Last annotation update)
 DE D1281H8.2 (PUTATIVE NOVEL PROTEIN SIMILAR TO KIAA0123 AND WORK
 DE C30F12.1) (FRAGMENT).
 GN D1281H8.2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Washieghl-Mohammadi M.;
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL031133; CAA20017.1; -
 FT NON-TER 1 160
 FT NON-TER 1 160
 SQ SEQUENCE 160 AA: 18983 MW: EC9F622F6CF4EF3 CRC64;

Query Match 90.3% Score 28; DB 4; Length 160;
 Best Local Similarity 66.7% Pred. No. 2.6e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXMR 6
 DB 30 FVPMWR 35

RESULT 9
 Q9BQZ9 PRELIMINARY: PRT: 160 AA.

AC 09BQZ9: 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE D1423H22.1 (NOVEL PROTEIN SIMILAR TO KIAA0123, KIAA0615 AND C.ELEGANS
 DE C30F12.1) (FRAGMENT).
 GN D1423H22.1 (FRAGMENT).

CN DJ423822.1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RX NCBL_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Pearce A.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL034379; CAC36083.1;
 FT NON_TER 1 160 160
 SQ SEQUENCE 160 AA: 18953 MW: 53EFA2B52DAED6C0 CRC64:

Query Match
 Best Local Similarity 90.3%; Score 28; DB 4; Length 160;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXWR 6
 DB 30 FVPSWR 35

RESULT 10
 ID 033867 PRELIMINARY; PRT: 211 AA.
 AC 033867;
 DT 01-JAN-1998 (TEMBLrel. 05, Created)
 DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
 DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
 DE SIMILAR TO ANSR GENBANK ACCESSION NUMBER 007683.
 OS Bacillus pumilus (Bacillus mesentericus).
 OC Plasmid pSH1452.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 RX NCBL_TaxID=1408;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Thomas C.M., Hasnain S.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U53767; AAB71491.1;
 DR InterPro: IPR001387; HTH_3.
 DR Pfam: PF01381; HTH_3; 1.
 DR SMART: SM00530; HTH_XRE; 1.
 KW Plasmid.
 SQ SEQUENCE 211 AA: 24620 MW: 49BD5954BD35E83A CRC64:

Query Match
 Best Local Similarity 90.3%; Score 28; DB 2; Length 211;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXWR 6
 DB 199 FIPTWR 204

RESULT 11
 ID 095726 PRELIMINARY; PRT: 218 AA.
 AC 095726;
 DT 01-MAY-1999 (TEMBLrel. 10, Created)
 DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TEMBLrel. 10, Last annotation update)
 DE HYPOTHETICAL 25.5 KDA PROTEIN (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RX NCBL_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RC TISSUE-BRAIN;
 RA Park S.H., Kim C., Kang Y.H., Chung H.S., Kim H.;
 RT "Expression of novel gene mapped on chromosome 7q22."
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U67037; AAO09521.1;
 KW Hypothetical protein.
 FT NON_TER 1 218 218
 SQ SEQUENCE 218 AA: 25458 MW: C0896F5EDAC5627F CRC64:

Query Match
 Best Local Similarity 90.3%; Score 28; DB 4; Length 218;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXWR 6
 DB 202 FVPSWR 207

RESULT 12
 ID 09XJ12 PRELIMINARY; PRT: 221 AA.
 AC 09XJ12;
 DT 01-NOV-1999 (TEMBLrel. 12, Created)
 DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)
 DT 01-JUN-2000 (TEMBLrel. 14, Last annotation update)
 DE HYPOTHETICAL PROTEIN.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzae; Oryza.
 RX NCBL_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. NIPPONBARE;
 RA Sasaki T., Nagamura Y., Yamamoto K.;
 RT clone: p0680A03."
 RT Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB023482; BAA78741.1;
 SQ SEQUENCE 221 AA: 24502 MW: DC24FF370660BC15 CRC64:

Query Match
 Best Local Similarity 90.3%; Score 28; DB 10; Length 221;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXWR 6
 DB 203 FRPTWR 208

RESULT 13
 ID 005115 PRELIMINARY; PRT: 236 AA.
 AC 005115;
 DT 01-JUL-1997 (TEMBLrel. 04, Created)
 DT 01-JUL-1997 (TEMBLrel. 04, Last sequence update)
 DT 01-NOV-1998 (TEMBLrel. 08, Last annotation update)
 DE PUTATIVE GLYCERATE KINASE AND PYRUVATE KINASE (PYKA) GENES, COMPLETE
 OS Methylobacterium extorquens.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Methylobacterium group; Methylobacterium.
 RX NCBL_TaxID=408;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AM1;
 RX MEDLINE=97386438; PubMed=9244287;
 RA Chistoserdova L., Lidstrom M.E.;
 RT "Identification and mutation of a gene required for glycerate kinase
 activity from a facultative methylotroph, Methylobacterium extorquens

RT AML":
RL J.Bacteriol. 179:4946-4948(1997).
DR EMBL: U87316; AAB66495.1; -
SQ SEQUENCE 236 AA; 26139 MW; 7F48376B148B3676 CRC64;

Query Match 90.3%; Score 28; DB 2; Length 236;
Best Local Similarity 66.7%; Pred. No. 3.7e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 FXPXMR 6
Db 135 FFPXMR 140

RESULT 14
ID 09VLD3 PRELIMINARY; PRT: 241 AA.
AC 09VLD3;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DE 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE CG3694 PROTEIN.
GN CG3694.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Gallie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton C.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agbayani A., An H.-J., Andrews-Plannkoch C., Baldwin D.,
RA Ballow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ilegam C.,
RA Jajali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matzel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacieb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spletter E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Sylvestre R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
DR EMBL: AE003624; AAF52759.1; -
DK FLYbase: FBgn003107; CG3694.

SO SEQUENCE 241 AA; 27071 MW; 200925FE35C3F733 CRC64;

Query Match 90.3%; Score 28; DB 5; Length 241;
Best Local Similarity 66.7%; Pred. No. 3.8e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 FXPXMR 6
Db 68 FFPVWR 73

RESULT 15
ID 09EUM2 PRELIMINARY; PRT: 250 AA.
AC 09EUM2;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE YCG4L.
GN YCG4L.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Plasmid R-plasmid pCG4.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;
OC Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC31830;
KA Tauch A., Puhler A., Kallowski J.;
RT "DNA sequence and genetic organization of the integron-carrying R-
RT plasmid pCG4 of Corynebacterium glutamicum";
KL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF164956; AAC00294.1; -
DR InterPro: IPR002145; COPG_HTM_4.
DR Pfam: PF01402; HTM_4; 1.
KW Plasmid.
SQ SEQUENCE 250 AA; 28804 MW; 195C863678483157 CRC64;

Query Match 90.3%; Score 28; DB 2; Length 250;
Best Local Similarity 66.7%; Pred. No. 4e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 FXPXMR 6
Db 214 FXPXMR 219

Search completed: February 27, 2002, 11:50:14
Job time: 993 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:36:07 ; Search time 132.19 Seconds
(without alignments)
1.021 Million cell updates/sec

Title: US-09-446-109a-17
Perfect score: 31
Sequence: 1 EXPXMR 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents_AA:*
1: /cgn2_6/prodata/2/iaa/5a_COMB.pep:*
2: /cgn2_6/prodata/2/iaa/5b_COMB.pep:*
3: /cgn2_6/prodata/2/iaa/6a_COMB.pep:*
4: /cgn2_6/prodata/2/iaa/6b_COMB.pep:*
5: /cgn2_6/prodata/2/iaa/PCTUS.COMB.pep:*
6: /cgn2_6/prodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	90.3	527	3	US-08-907-229-2
2	28	90.3	1536	4	US-09-413-814-10
3	28	90.3	2232	4	US-09-091-219-25
4	28	90.3	2247	4	US-09-091-219-2
5	27	87.1	6	1	US-08-215-137-13
6	27	87.1	456	1	US-09-268-364-21
7	26	83.9	42	2	US-08-766-858A-27
8	26	83.9	254	4	US-09-362-831-3
9	26	83.9	273	2	US-08-997-080-75
10	26	83.9	273	2	US-08-997-362-75
11	26	83.9	273	3	US-08-873-970-75
12	26	83.9	273	3	US-09-095-855-75
13	26	83.9	370	2	US-08-997-080-194
14	26	83.9	370	2	US-08-997-362-194
15	26	83.9	370	4	US-09-095-855-194
16	26	83.9	410	4	US-09-238-480-2
17	26	83.9	604	2	US-08-735-041A-4
18	26	83.9	604	2	US-09-190-476A-4
19	26	83.9	604	3	US-09-190-889A-4
20	26	83.9	604	4	US-09-190-938B-4
21	26	83.9	701	2	US-08-735-041A-6
22	26	83.9	701	3	US-09-190-476B-6
23	26	83.9	701	3	US-09-190-889A-6
24	26	83.9	701	4	US-09-190-938B-6
25	25	80.6	38	1	US-08-176-500-118
26	25	80.6	38	1	US-08-471-052A-118
27	25	80.6	38	1	US-08-189-331-118

28	25	80.6	38	2	US-08-471-939-118	Sequence 118, App
29	25	80.6	38	2	US-08-471-800-118	Sequence 118, App
30	25	80.6	38	2	US-08-471-068-118	Sequence 118, App
31	25	80.6	175	4	US-09-060-726A-2	Sequence 2, Appl1
32	25	80.6	631	1	US-08-487-890A-115	Sequence 115, App
33	25	80.6	631	2	US-08-478-435-115	Sequence 115, App
34	25	80.6	631	2	US-08-337-483-115	Sequence 115, App
35	25	80.6	631	2	US-08-478-373-115	Sequence 115, App
36	25	80.6	631	3	US-08-474-671-115	Sequence 115, App
37	25	80.6	631	3	US-08-483-577A-115	Sequence 115, App
38	25	80.6	631	4	US-08-897-438-115	Sequence 115, App
39	25	80.6	743	4	US-09-077-354B-2	Sequence 2, Appl1
40	24	77.4	10	3	US-08-159-339A-448	Sequence 448, App
41	24	77.4	15	2	US-08-553-257A-52	Sequence 52, Appl
42	24	77.4	175	1	US-08-414-926A-8	Sequence 8, Appl1
43	24	77.4	175	2	US-08-926-922-8	Sequence 8, Appl1
44	24	77.4	175	3	US-09-253-682-8	Sequence 8, Appl1
45	24	77.4	175	4	US-09-527-657-8	Sequence 8, Appl1

ALIGNMENTS

RESULT 1
US-08-907-229-2
Sequence 2, Application US/08907229A
Patent No. 6072048
GENERAL INFORMATION:
APPLICANT: Riley, Lee W.
TITLE OF INVENTION: DNA MOLECULE ENCODING FOR CELLULAR UPTAKE OF
FILE REFERENCE: 19603/1531
CURRENT APPLICATION NUMBER: US/08/907,229A
CURRENT FILING DATE: 1997-08-06
EARLIER APPLICATION NUMBER: 60/040,097
EARLIER FILING DATE: 1997-03-10
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 527
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-08-907-229-2

Query Match 90.3% Score 28; DB 3; Length 527;
Best Local Similarity 66.7% Pred. No. 2.9e+02;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Caps 0;

QY 1 EXPXMR 6
1 1 1 1
DB 3 FGPSMR 8

RESULT 2
US-09-413-814-10
Sequence 10, Application US/09413814
Patent No. 6225064
GENERAL INFORMATION:
APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
APPLICANT: Bristol-Myers Squibb, Co.
APPLICANT: Beyer, Stefan
APPLICANT: Bloeker, Helmut
APPLICANT: Brandt, Petra
APPLICANT: Cino, Paul M
APPLICANT: Dougherty, Brian A
APPLICANT: Goldberger, Steven L
APPLICANT: Hoefle, Gerhard
APPLICANT: Mueller, Joachim
APPLICANT: Reichenbach, Hans
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
TITLE OF INVENTION: heteropolymers of polyketide compounds
FILE REFERENCE: PCT/US 99/23535

CURRENT APPLICATION NUMBER: US/09/413.814
CURRENT FILING DATE: 1999-10-07
EARLIER APPLICATION NUMBER: DE 198 46 493.2
EARLIER FILING DATE: 1998-10-09
NUMBER OF SEQ ID NOS: 107
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 10
LENGTH: 1536
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-413-814-10

Query Match 90.3%; Score 28; DB 4; Length 1536;
Best Local Similarity 66.7%; Pred. NO. 8.1e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
| | | |
Db 896 FLPXMR 901

RESULT 3
US-09-091-219-25
Sequence 25, Application US/09091219
Patent No. 6171592
GENERAL INFORMATION:
APPLICANT: STUDDERT, Michael J.
APPLICANT: CRAAB, Brendan S.
APPLICANT: FENG, Li
TITLE OF INVENTION: EQUINE RHINOVIRUS 1 PROTEINS
FILE REFERENCE: 040268/0151
CURRENT APPLICATION NUMBER: US/09/091.219
CURRENT FILING DATE: 1998-10-05
EARLIER APPLICATION NUMBER: PCT/AU96/00815
EARLIER FILING DATE: 1996-12-18
EARLIER APPLICATION NUMBER: AU PN7201
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 25
LENGTH: 2232
TYPE: PRT
ORGANISM: equine rhinovirus 1
US-09-091-219-25

Query Match 90.3%; Score 28; DB 4; Length 2232;
Best Local Similarity 66.7%; Pred. NO. 1.1e+03;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
| | | |
Db 2227 FVPTMR 2232

RESULT 4
US-09-091-219-2
Sequence 2, Application US/09091219
Patent No. 6171592
GENERAL INFORMATION:
APPLICANT: STUDDERT, Michael J.
APPLICANT: CRAAB, Brendan S.
APPLICANT: FENG, Li
TITLE OF INVENTION: EQUINE RHINOVIRUS 1 PROTEINS
FILE REFERENCE: 040268/0151
CURRENT APPLICATION NUMBER: US/09/091.219
CURRENT FILING DATE: 1998-10-05
EARLIER APPLICATION NUMBER: PCT/AU96/00815
EARLIER FILING DATE: 1996-12-18
EARLIER APPLICATION NUMBER: AU PN7201
NUMBER OF SEQ ID NOS: 25

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 2247
TYPE: PRT
ORGANISM: equine rhinovirus 1
US-09-091-219-2

Query Match 90.3%; Score 28; DB 4; Length 2247;
Best Local Similarity 66.7%; Pred. NO. 1.2e+03;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
| | | |
Db 2227 FVPTMR 2232

RESULT 5
US-08-215-137-13
Sequence 13, Application US/08215137
Patent No. 5614370
GENERAL INFORMATION:
APPLICANT: Konteatis, Zenon
APPLICANT: Siciliano, Salvatore J
APPLICANT: Springer, Martin S
TITLE OF INVENTION: ASSAY TO IDENTIFY HUMAN C5A ANTAGONISTS
TITLE OF INVENTION: AND AGONISTS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: 126 Lincoln Ave., P.O. Box 2000
CITY: Rahway
STATE: New Jersey
COUNTRY: United States of America
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/215.137
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bencen, Gerard H
REGISTRATION NUMBER: 35,746
REFERENCE/DOCKET NUMBER: 19108
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-3901
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 6 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYDROPHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /label= BH
OTHER INFORMATION: /note= "either the natural phenylalanine amino
terminus or the Bolton-Hunter modified peptide
OTHER INFORMATION: having the group 3-(p-hydroxyphenyl)propionyl group
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /label= dcha
OTHER INFORMATION: /note= "D-cyclohexylalanine"

FEATURE:
NAME/KEY: Modified-site
LOCATION: 6
OTHER INFORMATION: /label= dvrg
OTHER INFORMATION: /note= "D-argline"
US-08-215-137-13

Query Match 87.1% Score 27; DB 1; Length 6;
Best Local Similarity 83.3% Pred. No. 1.6e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 1 FXPXMR-6
1 1111
Db 1 FXPXMR 6

RESULT 6
US-09-268-364-21
Sequence 21, Application US/09268364A
Patent No. 6204063
GENERAL INFORMATION:
APPLICANT: Allen, Steve
APPLICANT: Lee, Jian Ming
APPLICANT: Lightner, Jonathan
APPLICANT: Odell, Joan
TITLE OF INVENTION: PLANT GLYCOLYSIS AND RESPIRATION ENZYMES
FILE REFERENCE: BB-1154
CURRENT APPLICATION NUMBER: US/09/268,364A
CURRENT FILING DATE: 1999-03-15
EARLIER APPLICATION NUMBER: 60/079,387
EARLIER FILING DATE: March 16, 1998
NUMBER OF SEQ ID NOS: 21
SOFTWARE: Microsoft Office 97
SEQ ID NO 21
LENGTH: 456
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-268-364-21

Query Match 87.1% Score 27; DB 4; Length 456;
Best Local Similarity 66.7% Pred. No. 3.9e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Caps 0;

QY 1 FXPXMR 6
1 1 11
Db 209 FCPXMR 214

RESULT 7
US-08-766-858A-27
Sequence 27, Application US/08766858A
Patent No. 5935782
GENERAL INFORMATION:
APPLICANT: Cianciotto, Nicholas P.
APPLICANT: Hickey, Erin K.
APPLICANT: O'Connell, William A.
TITLE OF INVENTION: METHODS AND MATERIALS FOR DETECTING
TITLE OF INVENTION: LECTIONELLA PNEUMOPHILA
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross P.C.
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203

COMPUTER READABLE FORM:
MEDIUM TYPE: IBM PC compatible
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/766,858A
FILING DATE: 13-DEC-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/011,545
FILING DATE: 13-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Crook, Wannell M.
REGISTRATION NUMBER: 31,071
REFERENCE/DOCKET NUMBER: 3501-4-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 42 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-766-858A-27

Query Match 83.9% Score 26; DB 2; Length 42;
Best Local Similarity 50.0% Pred. No. 64;
Matches 3; Conservative 1; Mismatches 2; Indels 0; Caps 0;

QY 1 FXPXMR 6
1 1 1
Db 20 FSPXMR 25

RESULT 8
US-09-362-831-3
Sequence 3, Application US/09362831
Patent No. 6306400
GENERAL INFORMATION:
APPLICANT: BOBLIT et al.
TITLE OF INVENTION: AVIAN RECOMBINANT LIVE VACCINE USING, AS VECTOR, THE
TITLE OF INVENTION: AVIAN INFECTIOUS LARYNGOTRACHEITIS VIRUS
FILE REFERENCE: 454313-2520
CURRENT APPLICATION NUMBER: US/09/362,831
CURRENT FILING DATE: 1999-07-28
NUMBER OF SEQ ID NOS: 28
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 25
TYPE: PRT
ORGANISM: Infectious Laryngotracheitis Virus
US-09-362-831-3

Query Match 83.9% Score 26; DB 4; Length 254;
Best Local Similarity 66.7% Pred. No. 3.5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Caps 0;

QY 1 FXPXMR 6
1 1 11
Db 186 FCPXMR 191

RESULT 9
US-08-997-080-75
Sequence 75, Application US/08997080
Patent No. 5968524
GENERAL INFORMATION:
APPLICANT: WATSON, JAMES D.
APPLICANT: TAN, PAUL L.J.
TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-
NUMBER OF SEQUENCES: 194
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman

STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/997,080
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000,1007
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 273 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-997-080-75

Query Match 83.98; Score 26; DB 2; Length 273;
Best Local Similarity 66.78; Pred. No. 3.8e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 EXPXMR 6
Db 267 FDPWMR 272

RESULT 10
US-08-997-362-75
Sequence 75, Application US/08997362
Patent No. 5985287
GENERAL INFORMATION:
APPLICANT: Tan, Paul
APPLICANT: Hiyyama, Jun
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Scott, Linda
APPLICANT: Prestidge, Ross
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
NUMBER OF SEQUENCES: 194
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/997,362
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970
FILING DATE: June 12, 1997
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347
FILING DATE: August 29, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000,1002c2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 273 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-997-362-75

Query Match 83.98; Score 26; DB 2; Length 273;
Best Local Similarity 66.78; Pred. No. 3.8e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 EXPXMR 6
Db 267 FDPWMR 272

RESULT 11
US-08-873-970-75
Sequence 75, Application US/08873970
Patent No. 6001361
GENERAL INFORMATION:
APPLICANT: Tan, Paul
APPLICANT: Hiyyama, Jun
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Scott, Linda
APPLICANT: Prestidge, Ross
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/873,970
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/705,347
FILING DATE: 29-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000,1002C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563

TELEX:
: INFORMATION FOR SEQ ID NO: 75:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 273 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-873-970-75

Query Match 83.9% Score 26; DB 3; Length 273;
Best Local Similarity 66.7%; Pred. No. 3.8e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EXPXMR 6
DB 267 FDPWMR 272

RESULT 12
US-09-095-855-75
: Sequence 75, Application US/09095855
: Patent No. 6160093
: GENERAL INFORMATION:
: APPLICANT: Tan, Paul
: APPLICANT: Visser, Elizabeth
: APPLICANT: Skinner, Margot
: APPLICANT: Prestidge, Ross
: TITLE OF INVENTION: Compounds and Methods for
: TITLE OF INVENTION: Treatment and Diagnosis of Mycobacterial Infections
: NUMBER OF SEQUENCES: 208
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Law Offices of Ann W. Speckman
: STREET: 2601 Elliott Avenue, Suite 4185
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98121
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/095,855
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/705,347
: FILING DATE: 29-AUG-1996
: APPLICATION NUMBER: 08/873,970
: FILING DATE: 12-JUN-1997
: APPLICATION NUMBER: 08/997,362
: FILING DATE: 23-DEC-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Sleath, Janet
: REGISTRATION NUMBER: 37,007
: REFERENCE/DOCKET NUMBER: 11000.1002c3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 206-269-0565
: TELEFAX: 206-269-0563
: TELEX:
: INFORMATION FOR SEQ ID NO: 75:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 273 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-09-095-855-75

Query Match 83.9% Score 26; DB 4; Length 273;
Best Local Similarity 66.7%; Pred. No. 3.8e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EXPXMR 6
DB 267 FDPWMR 272

RESULT 13
US-08-997-080-194
: Sequence 194, Application US/08997080
: Patent No. 5968524
: GENERAL INFORMATION:
: APPLICANT: WATSON, JAMES D.
: APPLICANT: TAN, PAUL L.J.
: TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-

NUMBER OF SEQUENCES: 194

CORRESPONDENCE ADDRESS:

ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/997,080

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Sleath, Janet

REGISTRATION NUMBER: 37,007

REFERENCE/DOCKET NUMBER: 11000.1007

TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-269-0565

TELEFAX: 206-269-0563

TELEX:

INFORMATION FOR SEQ ID NO: 194:

SEQUENCE CHARACTERISTICS:

LENGTH: 370 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-997-080-194

Query Match 83.9% Score 26; DB 2; Length 370;
Best Local Similarity 66.7%; Pred. No. 5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EXPXMR 6
DB 364 FDPWMR 369

RESULT 14
US-08-997-362-194
: Sequence 194, Application US/08997362
: Patent No. 5985287
: GENERAL INFORMATION:
: APPLICANT: Tan, Paul
: APPLICANT: Hiyama, Jun
: APPLICANT: Visser, Elizabeth
: APPLICANT: Skinner, Margot
: APPLICANT: Scott, Linda

APPLICANT: Prestidge, Ross
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
NUMBER OF SEQUENCES: 194
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/997,362
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. Patent Application NO. 5985287 08/873,970
FILING DATE: June 12, 1997
APPLICATION NUMBER: U.S. Patent Application NO. 5985287 08/705,347
FILING DATE: August 29, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002c2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ ID NO: 194:
SEQUENCE CHARACTERISTICS:
LENGTH: 370 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-997-362-194

Query Match 83.9%; Score 26; DB 2; Length 370;
Best Local Similarity 66.7%; Pred. No. 5e+02; 2; Indels 0; Gaps 0;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
DB 364 FDPWMR 369

RESULT 15
US-09-095-855-194
Sequence 194, Application US/09095855
Patent No. 6160093
GENERAL INFORMATION:
APPLICANT: Tan, Paul
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Prestidge, Ross
TITLE OF INVENTION: Compounds and Methods for
TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
NUMBER OF SEQUENCES: 208
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/095,855
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/705,347
FILING DATE: 29-AUG-1996
APPLICATION NUMBER: 08/873,970
FILING DATE: 12-JUN-1997
APPLICATION NUMBER: 08/997,362
FILING DATE: 23-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002c3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ ID NO: 194:
SEQUENCE CHARACTERISTICS:
LENGTH: 370 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-095-855-194

Query Match 83.9%; Score 26; DB 4; Length 370;
Best Local Similarity 66.7%; Pred. No. 5e+02; 2; Indels 0; Gaps 0;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
DB 364 FDPWMR 369

Search completed: February 27, 2002, 11:36:08
Job time: 147 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:41:19 : Search time 303.5 Seconds
(without alignments)
1.464 Million cell updates/sec

Title: us-09-446-109a-18
Perfect score: 31
Sequence: 1 FXPXMR 6

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A_Geneseq_1101.*
2: /SIDSR/gcgdata/geneseq/geneseqp/AA1980.DAT.*
3: /SIDSR/gcgdata/geneseq/geneseqp/AA1981.DAT.*
4: /SIDSR/gcgdata/geneseq/geneseqp/AA1982.DAT.*
5: /SIDSR/gcgdata/geneseq/geneseqp/AA1983.DAT.*
6: /SIDSR/gcgdata/geneseq/geneseqp/AA1984.DAT.*
7: /SIDSR/gcgdata/geneseq/geneseqp/AA1985.DAT.*
8: /SIDSR/gcgdata/geneseq/geneseqp/AA1986.DAT.*
9: /SIDSR/gcgdata/geneseq/geneseqp/AA1987.DAT.*
10: /SIDSR/gcgdata/geneseq/geneseqp/AA1988.DAT.*
11: /SIDSR/gcgdata/geneseq/geneseqp/AA1989.DAT.*
12: /SIDSR/gcgdata/geneseq/geneseqp/AA1990.DAT.*
13: /SIDSR/gcgdata/geneseq/geneseqp/AA1991.DAT.*
14: /SIDSR/gcgdata/geneseq/geneseqp/AA1992.DAT.*
15: /SIDSR/gcgdata/geneseq/geneseqp/AA1993.DAT.*
16: /SIDSR/gcgdata/geneseq/geneseqp/AA1994.DAT.*
17: /SIDSR/gcgdata/geneseq/geneseqp/AA1995.DAT.*
18: /SIDSR/gcgdata/geneseq/geneseqp/AA1996.DAT.*
19: /SIDSR/gcgdata/geneseq/geneseqp/AA1997.DAT.*
20: /SIDSR/gcgdata/geneseq/geneseqp/AA1998.DAT.*
21: /SIDSR/gcgdata/geneseq/geneseqp/AA1999.DAT.*
22: /SIDSR/gcgdata/geneseq/geneseqp/AA2000.DAT.*
23: /SIDSR/gcgdata/geneseq/geneseqp/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	90.3	71	22	AA16826
2	28	90.3	71	22	AA16826
3	28	90.3	71	22	AA16826
4	28	90.3	71	22	AA16826
5	28	90.3	71	22	AA16826
6	28	90.3	71	22	AA16826
7	28	90.3	71	22	AA16826
8	28	90.3	71	22	AA16826
9	28	90.3	71	22	AA16826
10	28	90.3	71	22	AA16826
11	28	90.3	71	22	AA16826

12	28	90.3	629	20	AA135425
13	28	90.3	766	22	AAE03765
14	28	90.3	949	22	AAE95425
15	28	90.3	2247	18	AAW27126
16	27	87.1	6	13	AAE03577
17	27	87.1	6	16	AAE89737
18	27	87.1	30	22	AAE02765
19	27	87.1	31	22	AAE02764
20	27	87.1	39	22	AAW74031
21	27	87.1	79	22	AAW41295
22	27	87.1	312	22	AAE92505
23	27	87.1	359	22	AAE92586
24	27	87.1	392	21	AAE21559
25	27	87.1	399	21	AAE21558
26	27	87.1	408	21	AAE29695
27	27	87.1	434	21	AAE01208
28	27	87.1	434	21	AAE29694
29	27	87.1	449	20	AAE90188
30	27	87.1	449	21	AAE29693
31	27	87.1	499	21	AAE21054
32	27	87.1	588	21	AAE47006
33	27	87.1	594	21	AAE47005
34	27	87.1	704	22	AAE64416
35	27	87.1	719	22	AAE00798
36	27	87.1	724	22	AAE00911
37	27	87.1	761	20	AAE70899
38	27	87.1	761	21	AAE44704
39	27	87.1	846	21	AAE24241
40	27	87.1	897	22	AAE47046
41	27	87.1	1093	14	AAE41001
42	27	87.1	1204	21	AAE51233
43	26	83.9	160	22	AAE81828
44	26	83.9	254	19	AAE71200
45	26	83.9	262	21	AAE45062

ALIGNMENTS

RESULT 1

AA16826 standard; Protein: 71 AA.

AA16826;

12-OCT-2001 (first entry)

Peptide #3260 encoded by probe for measuring cervical gene expression.

Probe: human; microarray; gene expression; cervical epithelial cell;

cervical cancer.

Homo sapiens.

W0200157278-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00670.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632366.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000US-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI: 2001-488901/53.

Protein,
Human gene,
Human protein,
Equine rhinovirus,
Anaphylatoxin anta
C5a peptide analog
Circulin A synthet
Kallista B1 synthet
Human D2H binding
Human polypeptide
Human protein sequ
Human protein sequ
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Corn putative lect
Arabidopsis thalia
Arabidopsis Golgi
Arabidopsis thalia
Allialia isoflavone
Arabidopsis thalia
Arabidopsis thalia
Human nucleoprotein
Human bone marrow
Protein encoded by
Human tumor suppressor
Human KCNO5 (KCNQ
Human KCNO5 potass
Human myotonic dys
Newcastle disease
S. epidermidis ope
Protein encoded by
Arabidopsis thalia

XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 27; SEQ ID No 21652; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes
CC (SENPs: see A110068-A128459). The present sequence is a peptide encoded
CC by one such probe. The SENSs are derived from human HeLa cells. The SENSs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 71 AA:

Query Match 90.3%; Score 28; DB 22; Length 71;
Best Local Similarity 66.7%; Pred. No. 77;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXR 6
1 1 1 1
Db 1 fhpwr 6

RESULT 2
AAM29310
ID AAM29310 standard; Protein; 71 AA.

XX AAM29310:

XX 17-OCT-2001 (first entry)

XX Peptide #3347 encoded by probe for measuring placental gene expression.

XX Probe: microarray: human; placenta; antenatal diagnosis;

XX genetic disorder.

XX Homo sapiens.

XX MO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00663.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-48897/53.

XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 27; SEQ ID No 29579; 654pp; English.
XX The present invention relates to single exon nucleic acid probes (SENPs:
CC see A1131315-A157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for

CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
XX

XX Sequence 71 AA:

Query Match 90.3%; Score 28; DB 22; Length 71;
Best Local Similarity 66.7%; Pred. No. 77;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXR 6
1 1 1 1
Db 1 fhpwr 6

RESULT 3
AAM04538
ID AAM04538 standard; Protein; 71 AA.

XX AAM04538:

XX 09-OCT-2001 (first entry)

XX Peptide #3220 encoded by probe for measuring breast gene expression.

XX Probe: human; breast disease; breast cancer; development disorder;

XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX Homo sapiens.

XX MO200157270-A2.

XX 09-AUG-2001.

XX 29-JAN-2001; 2001WO-US00661.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-476286/51.

XX Novel single exon nucleic acid probe used to measuring gene expression
PT in a human breast -
XX
PS Claim 27; SEQ ID No 13278; 322pp; English.

XX The present invention relates to novel single exon nucleic acid probes
CC (see A1100010-A110067). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for measuring human gene expression in
CC a human breast sample, where the probe hybridises at high stringency to a
CC nucleic acid expressed in the human breast. The probes are useful for
CC predicting, diagnosing, grading, staging, monitoring and prognosing
CC diseases of the human breast, particularly those diseases with polygenic
CC aetiology. The diseases include: breast cancer, disorders of development,
CC inflammatory diseases of the breast, fibrocystic changes, proliferative
CC breast disease and non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 71 AA:

Query Match 90.3%; Score 28; DB 22; Length 71;
 Best Local Similarity 66.7%; Pred. No. 77;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 DY 1 FXPXWR 6
 | | | |
 DB 1 fhpawr 6

RESULT 4

AAB41139
 ID AAB41139; standard; Protein: 277 AA.

XX AAB41139;
 XX
 DT 08-FEB-2001 (first entry)

XX Human ORFX ORF903 polypeptide sequence SEQ ID NO:1806.

XX Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
 KM vulnery; antipapillary; antipapillary; antipapillary; antipapillary;
 KM anticonvulsant; osteopathic; antipapillary; antipapillary; antipapillary;
 KM immunostimulant; thrombotic; coagulant; vasotrophic; antidiabetic;
 KM hypotensive; dermatological; immunosuppressive; antihypertensive;
 KM antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;
 KM antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KM neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KM cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KM cholesterol ester storage; systemic lupus erythematosus; infection;
 KM severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KM allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KM bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KM thrombosis; contraceptive.

XX Homo sapiens.

XX MO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000MO-US08621.

XX 31-MAR-1999; 99US-0127607.

XX 02-APR-1999; 99US-0127636.

XX 05-APR-1999; 99US-0127728.

XX 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

XX Shimkels RA, Leach M;

XX WPI: 2000-602362/57.

XX N-PSDB: AAC75348.

XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -

XX Claim 11; Page 1407; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytosolic; hepatotropic; vulnery;
 CC antipapillary; antipapillary; antipapillary; antipapillary;
 CC osteopathic; anticonvulsant; antipapillary; antipapillary;
 CC immunostimulant; cardiac; thrombotic; coagulant; vasotrophic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antihypertensive; antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,

CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antihypertensive disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 277 AA;

Query Match 90.3%; Score 28; DB 21; Length 277;
 Best Local Similarity 66.7%; Pred. No. 2,9e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

DY 1 FXPXWR 6
 | | | |
 DB 271 fapwyr 276

RESULT 5

AAM24361
 ID AAM24361 standard; Protein: 357 AA.

XX AAM24361;

XX 12-OCT-2001 (first entry)

XX Human EST encoded protein SEQ ID NO: 1886.

XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;

XX tomato; monkey; dog; sea urchin; expressed sequence tag; EST;

XX diagnostics; forensic test; gene mapping; genetic disorder;

XX biodiversity; gene therapy; nutrition.

XX Homo sapiens.

XX MO200154477-A2.

XX 02-AUG-2001.

XX 25-JAN-2001; 2001MO-US02687.

XX 25-JAN-2000; 2000US-0491404.

XX 17-JUL-2000; 2000US-0617746.

XX 03-AUG-2000; 2000US-0631451.

XX 15-SEP-2000; 2000US-0663870.

XX (HYSE-) HYSE INC.

XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;

XX Cao Y, Drmanac RA, Zhang J, Werhman T;

XX WPI: 2001-476164/51.

XX N-PSDB: AAB99020.

XX Isolated polypeptide for treatment of diseases, diagnostics, raising
 PT antibodies and research use -

XX Claim 20; Page 1218-1219; 1275pp; English.

XX The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a
 CC protein of the invention.

XX Sequence 357 AA;

Query Match 90.3%; Score 28; DB 22; Length 357;
 Best Local Similarity 66.7%; Pred. No. 3.8e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 1 1 1 1
 241 fsprvr 246

RESULT 6
 ID AAU00428 standard; protein: 383 AA.
 AAU00428:

14-MAY-2001 (first entry)

Rat Gas1 protein.

Rat; Gas1; lethal cell protein; apoptosis; cell death;
 neurological disorder; cardiovascular disorder; autoimmune disorder;
 neuroendocrine disorder; oncological disorder; cancer.

Rattus norvegicus.

Key Location/Qualifiers
 Misc-difference 167
 /note="Encoded by TCC"
 Domain 174..279
 /note="Amphipathic alpha-helix region involved in
 channel activity and in death-inducing property
 of Gas1"

MO200114549-A1.

01-MAR-2001.

21-AUG-2000; 2000MO-EP08182.

24-AUG-1999; 99EP-0306702.

(JANC) JANSSEN PHARM NV.

Luyten WHML, Naranjo JR, Mellstroem B;

WPI: 2001-218449/22.

N-PSDB: AAU00428.

Inhibiting lethal effect of protein in cell, involves inhibiting
 function and/or expression of Gas1 protein and expression of sequence
 encoding otherwise lethal protein in the cell

Claim 32: Page 58-59; 65pp; English.

The present sequence represents rat Gas1 protein, a lethal cell protein.
 Novel methods of inhibiting the lethal effect of Gas1 or a functional
 equivalent, derivative or bioprecursor which is capable of inducing
 apoptosis (AAU00429) involve the use of antisense nucleotide sequences
 (AAS00433-AAS00434) which inhibit the expression of the lethal proteins
 in a cell. A nucleotide encoding Gas1 or a protein capable of inducing
 apoptosis, an antisense molecule, Gas1 polypeptide or fragment of, an
 inhibitor or accelerator of cell death, or a pharmaceutical composition
 comprising any of these are useful as a medicament for the prevention or
 treatment of a disease condition such as neurological disorder
 (Parkinson's disease, Alzheimer's disease, Huntington's disease,
 amyotrophic lateral sclerosis, or a neurological condition caused by
 thrombosis or cerebral trauma), cardiovascular disorder (heart attack),
 autoimmune disorder (multiple sclerosis), neuroendocrine disorder
 (necrosis of the pituitary gland), or oncological disorder (cancer),
 mediated at least in part by expression of Gas1 or its functional
 equivalent, derivative or bioprecursor capable of inducing apoptosis in
 a cell or a protein in the pathway of which Gas1 is a component. An

CC antisense molecule is useful as a probe, as a medicament or in the
 CC preparation of a drug for treating the above mentioned diseases.
 CC
 XX
 SQ Sequence 383 AA;

Query Match 90.3%; Score 28; DB 22; Length 383;
 Best Local Similarity 66.7%; Pred. No. 4e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 1 1 1 1
 138 fsprvr 143

RESULT 7
 ID AAY36300 standard; protein: 488 AA.
 AAY36300:

17-SEP-1999 (first entry)

Human secreted protein encoded by gene 77.

Human; secreted protein; cancer; tumour; developmental abnormality;
 foetal deficiency; blood disorder; immune system disorder; inflammation;
 autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
 schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;
 atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;
 digestive disorder; endocrine disorder; infection; AIDS.

Homo sapiens.

WO9931117-A1.

24-JUN-1999.

17-DEC-1998; 98WO-US27059.

19-DEC-1997; 97US-0068369.
 18-DEC-1997; 97US-0068006.
 18-DEC-1997; 97US-0068007.
 18-DEC-1997; 97US-0068008.
 18-DEC-1997; 97US-0068053.
 18-DEC-1997; 97US-0068054.
 18-DEC-1997; 97US-0068057.
 18-DEC-1997; 97US-0068064.
 18-DEC-1997; 97US-0070923.
 19-DEC-1997; 97US-0068169.
 19-DEC-1997; 97US-0068365.
 19-DEC-1997; 97US-0068367.
 19-DEC-1997; 97US-0068368.

(HUMA-) HUMAN GENOME SCI INC.

Carter KC, Duan RD, Feng P, Ferrie AM, Florence C;
 Florence K, Greene JM, Janat F, Kyaw H, Moore PA;
 Niu J, Rosen CA, Ruben SM, Shi Y, Soppet DR, Wei Y;
 Yu G;

WPI: 1999-418749/35.

N-PSDB: AAY97992.

New isolated human genes encoding secreted polypeptides

Claim 11: Page 387-389; 537pp; English.

AAY97916 to AAY98029 represent 110 isolated human secreted protein
 genes. AAY36224 to AAY36227 represent the secreted proteins encoded by
 the 110 human genes. The genes and their corresponding secreted
 polypeptides are useful for preventing, treating or ameliorating medical
 conditions, e.g. by protein or gene therapy. Also pathological conditions

CC can be diagnosed by determining the amount of the new polypeptides in a
CC sample or by determining the presence of mutations in the new genes.
CC Specific uses are described for each of the 110 genes, based on which
CC tissues they are most highly expressed in, and include developing
CC products for the diagnosis or treatment of cancer, tumours, developmental
CC abnormalities and foetal deficiencies, blood disorders, diseases of the
CC immune system, autoimmune diseases, inflammation, allergies, Alzheimer's
CC and cognitive disorders, schizophrenia, arthritis, asthma, psoriasis,
CC sepsis, skin disorders, atherosclerosis, diabetes, cardiovascular
CC disorders, kidney disorders, digestive/endocrine disorders, infections
CC and AIDS. The polypeptides are also useful for identifying their binding
CC partners. The sequences given in AAX97907 to AAX97915 and AAY36223 are
CC used in the exemplification of the present invention.

SO Sequence 488 AA;

Query Match

Best Local Similarity 90.3%; Score 28; DB 20; Length 488;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Caps 0;

OY 1 FXPXMR 6
DB 241 fsp1wr 246

RESULT 8

AAC92932 standard; Protein: 509 AA.

AC AAC92932;

DT 26-SEP-2001 (first entry)

DE C glutamicum protein fragment SEQ ID NO: 6686.

KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

KM organic acid synthesis.

OS Corynebacterium glutamicum.

PN EPI108790-A2.

PD 20-JUN-2001.

PF 18-DEC-2000; 2000EP-0127688.

PR 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX (KYOW) KYOWA HAKKO KOCYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI: 2001-376931/40.

DR N-PSDB: AAH68151.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying

PT mutation point of a gene, measuring expression of a gene, analysing

PT expression profile or pattern of a gene and identifying homologous gene

PT -

XX Claim 17: SEQ ID NO: 6686; 246bp + Sequence Listing: English.

XX The present invention provides a number of nucleotide and protein

CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These

CC are useful for identifying the mutation point of a gene derived from a

CC mutant of coryneform bacterium, measuring expression amount and

CC analysing the expression profile or expression pattern of a gene derived

CC from Coryneform bacterium, and identifying a homologue of a gene derived

CC from coryneform bacterium. Coryneform bacteria are useful for producing

CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.

SO Sequence 509 AA;

Query Match

Best Local Similarity 90.3%; Score 28; DB 22; Length 509;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Caps 0;

OY 1 FXPXMR 6
DB 364 fap1wr 369

RESULT 9

AAB79335 standard; Protein: 509 AA.

AC AAB79335;

DT 30-APR-2001 (first entry)

DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:186.

XX Corynebacterium glutamicum; carbon metabolism and energy production;

KW SMP protein; sugar metabolism and oxidative phosphorylation protein;

KW fine chemical production; organic acid; pyrimidine base; nucleoside;

KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;

KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;

KW carbonyl derivative; aromatic compound; vitamin; cofactor; polyketide; enzyme;

XX diagnosis; Corynebacterium diptheriae; evolutionary study.

OS Corynebacterium glutamicum.

PN WO200100844-A2.

PD 04-JAN-2001.

PF 23-JUN-2000; 2000WO-1B00943.

PR 25-JUN-1999; 99US-0141031.

PR 08-JUL-1999; 99DE-1031412.

PR 08-JUL-1999; 99DE-1031413.

PR 08-JUL-1999; 99DE-1031419.

PR 08-JUL-1999; 99DE-1031420.

PR 08-JUL-1999; 99DE-1031424.

PR 08-JUL-1999; 99DE-1031428.

PR 08-JUL-1999; 99DE-1031431.

PR 08-JUL-1999; 99DE-1031433.

PR 08-JUL-1999; 99DE-1031434.

PR 08-JUL-1999; 99DE-1031510.

PR 08-JUL-1999; 99DE-1031562.

PR 08-JUL-1999; 99DE-1031634.

PR 09-JUL-1999; 99DE-1032227.

PR 09-JUL-1999; 99DE-1032230.

PR 09-JUL-1999; 99US-0143208.

PR 14-JUL-1999; 99DE-1032924.

PR 14-JUL-1999; 99DE-1032973.

PR 14-JUL-1999; 99DE-1033005.

PR 27-AUG-1999; 99US-0140765.

PR 31-AUG-1999; 99US-0151572.

PR 03-SEP-1999; 99DE-1042076.

PR 03-SEP-1999; 99DE-1042079.

PR 03-SEP-1999; 99DE-1042086.

PR 03-SEP-1999; 99DE-1042087.

PR 03-SEP-1999; 99DE-1042088.

PR 03-SEP-1999; 99DE-1042095.

```

PR 03-SEP-1999; 99DE-1042123.
PR 03-SEP-1999; 99DE-1042125.
XX (BADI ) BASF AG.
XX
PI Pompeius M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;
XX
DR WPI: 2001-061975/07.
DR N-PSDB: AAF71452.
XX
PT New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
PT metabolism and oxidative phosphorylation protein for production or
PT modulation of production of fine chemicals e.g. amino acids,
PT carbohydrates or enzymes -
XX
PS Claim 20; Page 407-409; 1246pp; English.
XX
CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
CC metabolism and oxidative phosphorylation (SMP) proteins given in
CC AAF79243 to AAF79633 which are involved in carbon metabolism and
CC energy production. The C. glutamicum SMP gene can be used in vectors
CC (II) for expression in host cells and production or modulation of
CC production of fine chemicals, such as, an organic acid, a proteogenic
CC or nonproteogenic amino acid (preferred), a purine or pyrimidine base,
CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
CC cofactor, a polypeptide, or an enzyme. The presence of (I) or SMP proteins
CC (III) encoded by them are used for diagnosing the presence or activity of
CC Corynebacterium diphtheriae in a subject. (I), (II), (III) of host cells
CC containing them are used to map genomes of organisms related to
CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
CC in evolutionary studies, in determining SMP protein regions required
CC for function, in modulating SMP protein activity, in modulating the
CC metabolism of sugars, and in modulating high-energy molecule production
CC in a cell (i.e. ATP, NADPH).
XX
SQ Sequence 509 AA:

```

```

Query Match 90.3%; Score 28; DB 22; Length 509;
Best Local Similarity 66.7%; Pred. No. 5.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Caps 0;

```

```

QY 1 FXPXMR 6
   1 1 1 1
Db 364 fapfwr 369

```

```

RESULT 10
AAW5776
ID AAW5776 standard; Protein: 527 AA.
XX
AC AAW5776;
XX
DT 21-DEC-1998 (first entry)
XX
DE Mycobacterium tuberculosis cellular uptake Mcep protein.
XX
KW Mycobacterium cell entry; Mcep; cellular uptake; tuberculosis;
KW infection; vaccine; therapy; diagnosis; antibody; macrophage.
XX
OS Mycobacterium tuberculosis strain H37Ra (ATCC 25177).
XX
XX
XX Key Location/Qualifiers
XX FT MISC-difference 66 /note- "encoded by AT"
XX FT MISC-difference 202 /note- "encoded by CG"
XX
XX MO9840098-A1.
XX PD 17-SEP-1998.
XX

```

```

PF 09-MAR-1998; 98WO-US04653.
XX
XX 06-AUG-1997; 97US-0907229.
PR 10-MAR-1997; 97US-0040097.
XX
XX (CORR ) CORNELL RES FOUND INC.
XX
XX Riley LW;
XX
DR WPI: 1998-506483/43.
DR N-PSDB: AAV52613.
XX
XX
XX New isolated protein from Mycobacterium tuberculosis - useful for
XX e.g. diagnosis, treatment or prevention of infection and for
XX mediating uptake of materials by mammalian cells
XX
PS Claim 9; Page 12-13; 62pp; English.
XX
XX This is the deduced amino acid sequence of a protein encoded by a
XX DNA molecule (see AAV52613) associated with conferring on
XX Mycobacterium tuberculosis an ability to enter mammalian cells.
XX The protein, termed Mcep, or its fragments, especially polypeptides
XX comprising amino acids 1-167, 95-167, 95-528 or 379-528 of the
XX protein, can be expressed in host cells for recombinant production.
XX The proteins are useful in vaccines to prevent infection by M.
XX tuberculosis, and antibodies raised against the proteins can be
XX used for passive immunisation. The proteins, antibodies and DNA
XX molecules may be utilised in diagnostic assays to detect M.
XX tuberculosis in tissue or bodily fluids. The protein can be
XX associated with various other therapeutic materials (e.g.
XX antibiotics, DNA fragments or anti-neoplastic agents) for
XX administration to mammals, particularly humans, to achieve cellular
XX uptake of those materials. The ability of Mcep to augment uptake
XX into macrophages provides an opportunity to deliver genes
XX specifically to macrophages to induce humoral and cell-mediated
XX immunity.
XX
SQ Sequence 527 AA:

```

```

Query Match 90.3%; Score 28; DB 19; Length 527;
Best Local Similarity 66.7%; Pred. No. 5.5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Caps 0;

```

```

QY 1 FXPXMR 6
   1 1 1 1
Db 3 fgpwr 8

```

```

RESULT 11
AAW48359
ID AAW48359 standard; Protein: 527 AA.
XX
AC AAW48359;
XX
DT 03-JUL-1998 (first entry)
XX
DE Mycobacterium bovis BCG protein.
XX
KW Mycobacterium bovis BCG protein; Mcep; tuberculosis; BCGINV-33;
KW hybridisation test; amplification; probe.
XX
OS Mycobacterium bovis.
XX
XX
XX MO9801559-A1.
XX PD 15-JAN-1998.
XX
XX 09-JUL-1997; 97WO-CA00484.
XX PD 10-JUL-1996; 96US-0677970.
XX
XX (CONN-) CONNUGHT LAB LTD.
XX

```

Query Match	Best Local Similarity	Score 28:	DB 20:	Length 629:								
Matches 4:	Conservative	0:	Mismatches 2:	Indels 0: Gaps 0:								
Cy 1	FXPMXR 6											
Db 290	laplwr 295											
<p>RESULT 13</p> <p>AAE03765</p> <p>ID AAE03765 standard: Protein; 766 AA.</p> <p>AAE03765:</p> <p>07-AUG-2001 (first entry)</p> <p>Human gene 2 encoded secreted protein HCE3C63, SEQ ID NO:35.</p> <p>Human: secreted protein; proliferative disorder; cancer; tumour; foetal abnormality; developmental abnormality; haematopoietic disorder; immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis; inflammation; allergy; neurological disorder; Alzheimer's disease; Parkinson's disease; cognitive disorder; schizophrenia; asthma; skin disorder; psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder; angiotensin disorder; kidney disorder; gastrointestinal disorder; pregnancy-related disorder; gene therapy; endocrine disorder; infection; wound healing; veterinary; cell culture; chemotaxis; food additive; binding partner identification.</p> <p>Homo sapiens.</p> <table border="1"> <thead> <tr> <th>Key</th> <th>Location/Qualifiers</th> </tr> </thead> <tbody> <tr> <td>Peptide</td> <td>1..33</td> </tr> <tr> <td>Protein</td> <td>/label= signal_peptide 34..766</td> </tr> <tr> <td></td> <td>/note= "Mature secreted protein"</td> </tr> </tbody> </table> <p>WO200132877.1.</p> <p>10-MAY-2001.</p> <p>17-OCT-2000: 2000WO-US28664.</p> <p>02-NOV-1999: 99US-0163085.</p> <p>17-DEC-1999: 99US-0172411.</p> <p>(HUMA-) HUMAN GENOME SCI INC.</p> <p>Fiscella M, Wei P, Lafleur DW, Olsen HS, Baker K, Ebner R; Komatsu S, G, Rosen CA, Ruben SM, Duan RP, Young PE, Florence KA; Moore PA, Birse CE, Ni J, Soppet DR, Shi Y; WPI: 2001-328782/34.</p> <p>N-PSDB: AAD08192.</p>					Key	Location/Qualifiers	Peptide	1..33	Protein	/label= signal_peptide 34..766		/note= "Mature secreted protein"
Key	Location/Qualifiers											
Peptide	1..33											
Protein	/label= signal_peptide 34..766											
	/note= "Mature secreted protein"											

PT Novel human secreted proteins and nucleic acids for diagnosing,
 PT preventing and treating neurological, cardiovascular, infectious,
 PT autoimmune, gastrointestinal, bone disorders, cancer, particularly
 PT ovarian cancer -
 PS Claim 11: Page 387-390; 421pp: English.

XX AAD08191-AAD08213 represent cDNAs corresponding to 19 human secreted
 CC protein genes and AAE03764-AAE03786 represent the proteins they encode.
 CC AAE03787-AAE03800 represent human secreted protein fragments or variants.
 CC The genes and their secreted proteins are useful for preventing,
 CC treating or ameliorating medical conditions, e.g., by protein or gene
 CC therapy. Pathological conditions can be diagnosed by determining the
 CC amount of the new protein in a sample or by determining the presence of
 CC mutations in the new genes. Specific uses are described for each of the
 CC 19 genes, based on the tissues in which they are most highly expressed,
 CC and include developing products for the diagnosis or treatment of
 CC proliferative disorders, cancer, tumours, foetal and developmental
 CC abnormalities, hematopoietic disorders, diseases of the immune system,
 CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
 CC allergies, neurological disorders (e.g., Alzheimer's disease,
 CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
 CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
 CC cardiovascular disorders, angiogenic disorders, kidney disorders,
 CC gastrointestinal disorders, pregnancy-related disorders, endocrine
 CC disorders, and infections. The proteins can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues, to identify their
 CC cognate ligands or binding partners, and in chemotaxis, and can be used
 CC as a food additive or preservative to modify storage properties.
 CC Antibodies specific for a protein of the invention can be used in
 CC allelic screening for a protein of the invention mentioned above, and
 CC in diagnostic assays associated with the disorders mentioned above, and
 CC in diagnostic immunoassays (e.g., radioimmunoassay or enzyme linked
 CC immunosorbent assay (ELISA)). The present sequence represents a human
 CC secreted protein of the invention.
 XX Sequence 766 AA:

Query Match 90.3%; Score 28; DB 22; Length 766;
 Best Local Similarity 66.7%; Pred. NO. 8e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXWR 6
 1 1 1 1
 DB 525 fdpsvr 530

RESULT 14
 AAB95425
 ID AAB95425 standard; Protein: 949 AA.

XX AAB95425:

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO: 17833.

XX Human: primer: detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Salto K, Yamamoto J;

XX Iishi S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI: 2001-318749/34.

XX primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -

XX Claim 8; SEQ ID 17833; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dt primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to

CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.

XX Sequence 949 AA:

Query Match 90.3%; Score 28; DB 22; Length 949;
 Best Local Similarity 66.7%; Pred. NO. 9.8e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXWR 6
 1 1 1 1
 DB 241 fspivr 246

RESULT 15
 AAM27126
 ID AAM27126 standard; Protein: 2247 AA.

XX AAM27126:

XX 11-FEB-1998 (first entry)

XX Equine rhinovirus 1 (ERHV1) polypeptide amino acid sequence.

XX Equine rhinovirus 1; ERHV1: foot-and-mouth disease virus; vaccine; horse;

XX diagnosis; antigens; polypeptide; enzyme-linked immunosorbent assay;

XX recombinant protein.

XX Equine rhinovirus 1.

XX Key Location/Qualifiers

XX Cleavage-site 207..208

XX Cleavage-site 287..288

XX Cleavage-site 517..518

XX Cleavage-site 743..744

XX Cleavage-site 991..992

W09722701-A1

26-JUN-1997

18-DEC-1996; 96WO-AU00815

18-DEC-1995; 95AU-0007201

(UYME) UNIV MELBOURNE

Crabb BS, Feng L, Studdert MJ

WPI; 1997-341692/31

N-PSDB; AAT85178

Genomic sequence of equine rhinovirus 1 - and derived proteins on virus-like particles, useful in vaccines and as diagnostic agents

Claim 2; Pages 30-32; 60pp; English

Sequence 2247 AA

```

Query Match          90.3%  Score 28:  DB 18:  Length 2247:
Best Local Similarity 66.7%  Pred. No. 2.3e+03:
Matches      4:  Conservative      0:  Mismatches      2:  Indels      0:  Gaps      0:
QY           1  FXPXWR 6
              | | | |
FbB  2227  fvpLwr  2232

Search completed: February 27, 2002, 11:41:20
Job time: 459 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:45:23 ; Search time 145.23 Seconds
(without alignments)
3.147 Million cell updates/sec

Title: US-09-446-109A-18

Perfect score: 31
Sequence: 1 FXPXMR 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

PIR_68: *
1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	93.5	354	2	T39599 conserved hypothet
2	29	93.5	477	2	S77373 hypothetical prote
3	29	93.5	563	3	S78224 virulence-associat
4	29	93.5	639	1	MMVQ70 70K protein - puta
5	28	90.3	135	2	E82578 hypothetical prote
6	28	90.3	233	1	VMMJBV envelope protein E
7	28	90.3	254	2	S10929 transposase - Myco
8	28	90.3	254	2	A49895 transposase - Flav
9	28	90.3	297	2	E84788 hypothetical prote
10	28	90.3	311	2	T32776 hypothetical prote
11	28	90.3	337	2	B84335 conserved hypothet
12	28	90.3	389	2	G83317 hypothetical prote
13	28	90.3	410	2	S18157 conserved hypothet
14	28	90.3	475	1	A69149 globulin 2 - easle
15	28	90.3	488	2	S18156 O-antigen transpo
16	28	90.3	517	2	A70793 globulin 1 - easle
17	28	90.3	517	2	A70793 probable glycerol
18	28	90.3	552	2	C70567 probable ilvB pro
19	28	90.3	635	2	G86589 hypothetical prote
20	28	90.3	635	2	G72035 regulatory protein
21	28	90.3	639	2	S03547 hypothetical prote
22	28	90.3	703	2	T48019 hypothetical prote
23	28	90.3	1308	2	B75198 DNA helicase relat
24	27	87.1	89	2	D84256 conserved hypothet
25	27	87.1	113	2	G82517 conserved hypothet
26	27	87.1	119	2	T36433 hypothetical prote
27	27	87.1	199	2	T22304 conserved hypothet
28	27	87.1	236	2	D72355 conserved hypothet
29	27	87.1	252	2	C70380 hypothetical prote

30	27	87.1	267	2	A75339 hypothetical prote
31	27	87.1	301	2	T24993 hypothetical prote
32	27	87.1	303	2	T46416 hypothetical prote
33	27	87.1	309	2	S75505 hypothetical prote
34	27	87.1	310	2	F86737 hypothetical prote
35	27	87.1	318	2	G87350 Lipid A biosynthes
36	27	87.1	340	2	C75531 cytochrome c fam11
37	27	87.1	345	2	T14707 DNA ligase homolog
38	27	87.1	365	2	T15010 hypothetical prote
39	27	87.1	380	2	C64364 formate hydrogenu
40	27	87.1	395	2	D81146 aminotransferase,
41	27	87.1	395	2	E81877 probable aminotran
42	27	87.1	412	2	E65146 conserved hypothet
43	27	87.1	422	2	C82666 conserved hypothet
44	27	87.1	449	2	C84618 hypothetical prote
45	27	87.1	456	2	S61170 BCS1 protein precu

ALIGNMENTS

RESULT 1
T39599
conserved hypothetical protein SPBC16G5.07c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 18-Feb-2000
C:Accession: T39599
R:Lyne, M.; Wood, V.; Rajandream, M.A.; Barrell, B.G.; Beck, A.; Reinhardt, R.
submitted to the EMBL Data Library, May 1998
A:Reference number: 221866
A:Accession: T39599
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-354 <LVN>
A:Cross-references: EMBL:AL023554; PIRN:CAA19027.1; GSPDB:GN00067; SPDB:SPBC16G5.07c
A:Experimental source: strain 972H-; cosmid c16G5
C:Genetics:
A:Gene: SPDB:SPBC16G5.07c
A:Map position: 2
A:Introns: 72/2: 265/2
C:Superfamily: erythrocyte band 7 integral membrane protein

Query Match 93.5% Score 29; DB 2; Length 354;
Best Local Similarity 66.7% Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXMR 6
DB 40 FXPXMR 45

RESULT 2
S77373
hypothetical protein s111464 - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C:Accession: S77373
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Miyajima, O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasu
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A:Reference number: S74322; MUID:97061201
A:Accession: S77373
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-477 <KAN>
A:Cross-references: EMBL:D90906; GB:AB001339; NID:q1652492; PIRN:BAI17476.1; PIR:dl01
C:Genetics:
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
A:Start codon: GTG

Query Match 93.5%; Score 29; DB 2; Length 477;
Best Local Similarity 66.7%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXMR 6
1 1 1 1
DB 388 FSPSWR 393

RESULT 3
S78224
virulence-associated protein mkfB - Salmonella typhimurium plasmid

C:Species: Salmonella typhimurium
C:Date: 29-Jan-1998 #sequence_revision 13-Feb-1998 #text_change 20-Aug-1999
C:Accession: S78224; S23713; A43996
R:Morel, F.

submitted to the EMBL Data Library, January 1992

A:Reference number: S78224

A:Accession: S78224

A:Molecule type: DNA

A:Residues: 1-563 <NOR>

A:Cross-references: EMBL:X57096; NID:g46998; PIDN:CAA40380.1; PID:g46999

R:Morel, F.; Pisano, M.R.; Niccoli, J.; Popoff, M.Y.

Res. Microbiol. 140, 455-457, 1989

A:Title: Nucleotide sequence of the plasmid-borne virulence gene mkfB from Salmonella ty

A:Reference number: A43996; MUID:90161559

A:Accession: S23713

A:Molecule type: DNA

A:Residues: 1-513; 'NILEKKGKRLGDVHFKGEAEMLF', 514-563 <NOM>

A:Cross-references: EMBL:X57096

C:Genetics:

A:Genome: plasmid

C:Superfamily: virulence-associated protein spvB

C:Keywords: DNA binding; transcription regulation

Query Match 93.5%; Score 29; DB 2; Length 563;
Best Local Similarity 66.7%; Pred. No. 2e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXMR 6
1 1 1 1
DB 50 FAPAWR 55

RESULT 4
WM070
70K protein - potato leaf roll virus (strain 1)

C:Species: potato leaf roll virus

A:Note: host Solanum tuberosum (potato)

C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 21-Jul-2000

C:Accession: JA0118; S24591

R:Mayo, M.A.; Robinson, D.J.; Jolly, C.A.; Hyman, L.

J. Gen. Virol. 70, 1037-1051, 1989

A:Title: Nucleotide sequence of potato leafroll luteovirus RNA.

A:Reference number: JA0119; MUID:89279282

A:Accession: JA0118

A:Molecule type: genomic RNA

A:Residues: 1-639 <NAV>

A:Cross-references: EMBL:X14600; NID:g222293; PIDN:BA00417.1; PID:g222297

A:Note: The nucleotide sequence was submitted to the EMBL Data Library, March 1989

C:Comment: The genome is a single-stranded, positive-sense RNA.

C:Superfamily: potato leaf roll virus 70K protein

Query Match 93.5%; Score 29; DB 1; Length 639;
Best Local Similarity 66.7%; Pred. No. 2.2e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXMR 6
1 1 1 1

DB 615 FFPSPWR 620

RESULT 5
E82578
hypothetical protein XF2271 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C:Accession: E82578

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717

A:Note: For a complete list of authors see reference number A59328 below

A:Accession: E82578

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-135 <SIM>

A:Cross-references: GB:AE004039; GB:AE003849; NID:g9107425; PIDN:AAE85070.1; GSPDB:GN

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Rehnach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.

Uriones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer

as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitejima, J.P.; Krieger, J.E.; Kuramae, E.E.; La

chado, M.A.; Madalena, A.M.B.N.; Madalena, H.M.F.; Martino, C.L.; Marques, M.V.; Martins

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.

, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeri,

Rodrigues, V.; Rosa, A.J.; de M.; de Rosa Jr., V.E.; de Sa, R.C.; Sanello, R.V.; Sava

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv

M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF2271

Query Match 90.3%; Score 28; DB 2; Length 135;
Best Local Similarity 66.7%; Pred. No. 82;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXMR 6
1 1 1 1
DB 79 FVPWWR 84

RESULT 6
VIMKJIV

envelope protein E - Berne virus

C:Species: Berne virus

C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 16-Jul-1999

C:Accession: A39989; S15570

R:Ben Boon, J.A.; Snijder, E.J.; Locker, J.K.; Horzinek, M.C.; Rottier, P.J.M.

Virol. 182, 655-663, 1991

A:Title: Another triple-spanning envelope protein among intracellularly budding RNA v

A:Reference number: A39989; MUID:91220715

A:Accession: A39989

A:Molecule type: genomic RNA

A:Residues: 1-233 <DEN>

A:Cross-references: GB:X52505; NID:958768; PIDN:CA36747.1; PID:958769

R:Boon den, J.A.; Snijder, E.J.; Horzinek, M.C.; Rottier, P.J.M.

submitted to the EMBL Data Library, February 1990

A:Reference number: S15570

A:Accession: S15570

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-233 <BOO>

A:Cross-references: EMBL:X52505; NID:958768; PIDN:CA36747.1; PID:958769

C:Superfamily: Berne virus envelope protein E

C:Keywords: envelope protein

Query Match
Best Local Similarity 90.3%; Score 28; DB 1; Length 233;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
DB 60 FSPXMR 65

RESULT 7

SI0929
Transposase - Mycobacterium fortuitum insertion sequence IS6100
C:Species: Mycobacterium fortuitum
A:Variety: strain FCI
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 24-Sep-1999
R:Accession: SI0929
R:Matin, C.; Tilm, J.; Raulier, J.; Gomez-Lus, R.; Davies, J.; Gicquel, B.
Nature 345, 739-743, 1990
A:Title: Transposition of an antibiotic resistance element in mycobacteria.
A:Reference number: SI0927; MUID:90294910
A:Accession: SI0929
A:Molecule type: DNA
A:Residues: 1-254 <NMT>
A:Cross-references: EMBL:X53635; NID:g44283; PIDN:CAA37685.1; PID:g44285
A:Experimental source: strain FCI
C:Genetics:
A:Mobile element: insertion sequence IS6100
C:Superfamily: Salmonella typhimurium conserved hypothetical protein
C:Keywords: DNA binding

Query Match
Best Local Similarity 90.3%; Score 28; DB 2; Length 254;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
DB 70 FDPXMR 75

RESULT 8

A49895
Transposase - Flavobacterium sp. insertion sequence IS6100
C:Species: Flavobacterium sp.
C:Date: 11-Aug-1995 #sequence_revision 11-Aug-1995 #text_change 04-Sep-1998
R:Accession: A49895
R:Kato, K.; Ohtsuki, K.; Mitsuoka, H.; Yomo, T.; Negoro, S.; Urahe, I.
J. Bacteriol. 176, 1197-1200, 1994
A:Title: Insertion sequence IS6100 on plasmid pOAD2, which degrades nylon oligomers.
A:Reference number: A49895; MUID:94148782
A:Accession: A49895
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-254 <KAT>
A:Cross-references: GB:D26094
A:Experimental source: strain K172371; plasmid pOAD2; insertion sequence IS6100
C:Genetics:
A:Note: nucleotide sequence not given
A:Mobile element: insertion sequence IS6100
C:Superfamily: Salmonella typhimurium conserved hypothetical protein
C:Keywords: DNA binding

Query Match
Best Local Similarity 90.3%; Score 28; DB 2; Length 254;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
DB 70 FDPXMR 75

RESULT 9

584788

hypothetical protein At2g37100 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
R:Accession: E84788
R:Lin, X.; Kaul, S.; Kounsley, S.D.; Shea, T.P.; Benito, M.L.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Mofatt, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
Guss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: E84788
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-297 <STO>
A:Cross-references: GB:AE002093; NID:g4371291; PIDN:AAD18149.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g37100
A:Map position: 2

Query Match
Best Local Similarity 90.3%; Score 28; DB 2; Length 297;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
DB 118 FSPXMR 123

RESULT 10

T32776
hypothetical protein D1069.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 04-Mar-2000
R:Accession: T32776
R:Murray, J.; Langston, Y.; Clarke, K.; Morris, M.
Submitted to the EMBL data library, December 1997
A:Description: The sequence of C. elegans cosmid D1069.
A:Reference number: 221223
A:Accession: T32776
A>Status: preliminary; translated from CU/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-311 <MUR>
A:Cross-references: EMBL:AF040641; PIDN:AA94946.1; GSPDB:GN000020; CESP:D1069.1
A:Experimental source: strain Bristol N2; clone D1069
C:Genetics:
A:Gene: CESP:D1069.1
A:Map position:
A:Intons: 264
C:Superfamily: Caenorhabditis elegans hypothetical protein D1069.1

Query Match
Best Local Similarity 90.3%; Score 28; DB 2; Length 311;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 5
DB 25 FFPXMR 30

RESULT 11

B84335
hypothetical protein Vng1839h [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
R:Accession: B84335
R:Ng, W.V.; Kennedy, S.P.; Mahafas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky,
J.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebdardt, H.; Lowe, T.M.;

A:Title: Genome sequence of Halobacterium species NRC-1.
 A:Reference number: AB4160; MUID:20504483
 A:Accession: B84335
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-337 <STO>
 A:Cross-references: GB:AE004437; NID:g10581284; PIDN:AAG20046.1; GSPDB:GN00138
 C:Genetics:
 A:Gene: VNG1839H

Query Match 90.3%; Score 28; DB 2; Length 337;
 Best Local Similarity 66.7%; Pred. No. 1.9e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPXMR 6
 1 1 1 1
 DB 79 FVPSMR 84

RESULT 12
 G83317
 Conserved hypothetical protein PA2630 [Imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: G83317
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B.
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim,
 ; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A:Reference number: AB2950; MUID:20437337
 A:Accession: G83317
 A:Molecule type: DNA
 A:Status: preliminary
 A:Residues: 1-389 <STO>
 A:Cross-references: GB:AE004691; GB:AE004091; NID:g9948688; PIDN:AAG06018.1; GSPDB:GN001
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA2630

Query Match 90.3%; Score 28; DB 2; Length 389;
 Best Local Similarity 66.7%; Pred. No. 2.2e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPXMR 6
 1 1 1 1
 DB 111 FVPSMR 116

RESULT 13
 S18157
 globulin 2 - eastern white pine
 C:Species: Pinus strobus (eastern white pine)
 C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Aug-1999
 C:Accession: S18157
 R:Rugh, C.L.; Kamalay, J.C.
 submitted to the EMBL Data Library, November 1991
 A:Description: Legumin mRNAs from Pinus strobus L.
 A:Reference number: S18156
 A:Accession: S18157
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-410 <RUG>
 A:Cross-references: EMBL:211487; NID:g20744; PIDN:CAA77569.1; PID:g20745
 C:Superfamily: glycinn

Query Match 90.3%; Score 28; DB 2; Length 410;
 Best Local Similarity 66.7%; Pred. No. 2.3e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPXMR 6
 1 1 1 1
 DB 269 FVPSMR 274

RESULT 14
 A69149
 O:antigen transporter related protein - Methanobacterium thermoautotrophicum (strain
 C:Species: Methanobacterium thermoautotrophicum
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
 C:Accession: A69149
 R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T
 ; Qiu, D.; Spadafora, R.; Vitale, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jivani,
 K.; S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
 J. Bacteriol. 179, 7135-7155, 1997
 A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
 A:Reference number: A69000; MUID:98037514
 A:Accession: A69149
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-475 <MTH>
 A:Cross-references: GB:AE000823; GB:AE000666; NID:g2621432; PIDN:AB84885.1; PID:g262
 A:Experimental source: strain Delta H
 C:Genetics:
 A:Gene: MTH379
 A:Start codon: GTG
 C:Superfamily: succinoglycan biosynthesis transport protein

Query Match 90.3%; Score 28; DB 1; Length 475;
 Best Local Similarity 66.7%; Pred. No. 2.6e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPXMR 6
 1 1 1 1
 DB 408 FVPSMR 413

RESULT 15
 S18156
 globulin 1 - eastern white pine
 C:Species: Pinus strobus (eastern white pine)
 C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Aug-1999
 C:Accession: S18156
 R:Rugh, C.L.; Kamalay, J.C.
 submitted to the EMBL Data Library, November 1991
 A:Description: Legumin mRNAs from Pinus strobus L.
 A:Reference number: S18156
 A:Accession: S18156
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-488 <RUG>
 A:Cross-references: EMBL:211486; NID:g20742; PIDN:CAA77568.1; PID:g20743
 C:Superfamily: glycinn

Query Match 90.3%; Score 28; DB 2; Length 488;
 Best Local Similarity 66.7%; Pred. No. 2.7e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPXMR 6
 1 1 1 1
 DB 347 FVPSMR 352

Search completed: February 27, 2002, 11:45:23
 Job time: 702 sec

Wed Feb 27 12:13:57 2002

us-09-446-109a-18.rpr

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:42:48 : Search time 78.39 seconds
(without alignments)

2.806 Million cell updates/sec

Title: US-09-446-109a-18
Sequence: 1 FXPXMR 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: SWISSPROT_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	93.5	452	1	SPS2_MOUSE
2	29	93.5	477	1	P7364 mus musculu
3	29	93.5	477	1	P7364 synchocyst
4	28	90.3	151	1	P70K_PLEVI3
5	28	90.3	151	1	HSPB_BRAJA
6	28	90.3	151	1	HSPB_BRAJA
7	28	90.3	153	1	HSPB_BRAJA
8	28	90.3	233	1	VENV_BEV
9	28	90.3	333	1	DPSD_CAEEL
10	28	90.3	508	1	GLPK_MYCLE
11	28	90.3	517	1	GLPK_MYCYU
12	27	87.1	199	1	VAF7_CAEEL
13	27	87.1	277	1	HO_FUGRU
14	27	87.1	456	1	BCSL_YEAST
15	27	87.1	459	1	RBL2_RHOSH
16	27	87.1	514	1	GSH1_BRAJU
17	27	87.1	522	1	GSH1_ARATH
18	27	87.1	535	1	VHIL_ECOLI
19	27	87.1	591	1	IF37_ARATH
20	27	87.1	623	1	PRR2_CANAL
21	27	87.1	644	1	VP4_BRV2A
22	27	87.1	878	1	CI05_MOUSE
23	27	87.1	897	1	CI05_HUMAN
24	27	87.1	1023	1	TSCQ_PSEAM
25	27	87.1	1204	1	RRLP_NDVB
26	26	83.9	424	1	RPLA_LACIA
27	26	83.9	407	1	YK67_MYCTU
28	26	83.9	486	1	LPLI_YARLI
29	26	83.9	575	1	IDOA_ECOLI
30	26	83.9	700	1	YHFK_ECOLI
31	26	83.9	746	1	TAGF_BACSU
32	26	83.9	855	1	YR29_YEAST
33	26	83.9	3206	1	POLC_PSHMV

34	25	80.6	284	1	YAT1_SYNP1	O05380 synchococ
35	25	80.6	307	1	CRTB_SYNP7	P37269 synchococ
36	25	80.6	493	1	YEB6_YEAST	P39997 saccharomyc
37	25	80.6	497	1	CPD6_HUMAN	P10635 homo sapien
38	25	80.6	497	1	CPDH_MACFA	Q29488 macaca fasc
39	25	80.6	497	1	CPD4_MACFA	O18992 callitrich
40	25	80.6	500	1	CPD4_RAT	P13108 rattus norv
41	25	80.6	500	1	CPDG_CAVPO	O64403 cavia porce
42	25	80.6	500	1	CPDI_RAT	O64680 rattus norv
43	25	80.6	523	1	NU4M_PROMI	Q37617 proteolueca
44	25	80.6	628	1	YAXN_RHISN	P5706 rhizobium s
45	25	80.6	743	1	ANAG_HUMAN	P54802 homo sapien

ALIGNMENTS

RESULT	ID	SPS2_MOUSE	STANDARD	PRT	452 AA.
AC	P97364:				
DT	15-JUL-1999 (Rel. 38, Created)				
DT	15-JUL-1999 (Rel. 38, Last sequence update)				
DT	30-MAY-2000 (Rel. 39, Last annotation update)				
DE	SELENIDE, WATER DIKINASE 2 (EC 2.7.9.3) (SELENOPHOSPHATE SYNTHETASE 2)				
DE	(SELENIUM DONOR PROTEIN 2).				
GN	SPS2.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
FP	SEQUENCE FROM N.A.				
KX	MEDLINE=96017645; PubMed=7588067;				
KA	Guimaraes M.J., Bazan J.F., Zlotnick A., Wiles M.V., Grimaldi J.C.,				
KA	Lee F., McClanahan T.;				
KT	"A new approach to the study of haematopoietic development in the				
KT	yolk sac and embryoid bodies.";				
HL	Development 121:3335-3346(1995).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
KX	MEDLINE=97140286; PubMed=8986768;				
KA	Guimaraes M.J., Peterson D., Vicari A., Cocks B.G., Copeland N.G.,				
KA	Gilbert D.J., Jenkins N.A., Ferrick D.A., Kastolein R., Bazan J.F.,				
KA	Zlotnick A.;				
RT	"Identification of a novel seld homolog from eukaryotes, bacteria,				
RT	and archaea: is there an autoregulatory mechanism in selenocysteine				
RT	metabolism?";				
HL	Proc. Natl. Acad. Sci. U.S.A. 93:15086-15091(1996).				
CC	- FUNCTION: SYNTHESIZES SELENOPHOSPHATE FROM SELENIDE AND ATP.				
CC	- CATALYTIC ACTIVITY: ATP + SELENIDE + H(2)O = AMP + SELENOPHOSPHATE				
CC	+ PHOSPHATE.				
CC	- COFACTOR: SELENOCYSTEINE. THE ACTIVE-SITE SELENOCYSTEINE IS				
CC	ENCODED BY THE OPAL CODON, UGA.				
CC	- SIMILARITY: BELONGS TO THE SELENOPHOSPHATE SYNTHETASE 1 FAMILY.				
CC	CLASS 1 SUBFAMILY.				
UC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL: U43285; AAC53024.1; -				
DR	MGD: MGI:108388; SPS2.				
DR	Interpro: IPR000728; ATRS-related.				
DR	Pfam: PF00586; ATRS. 1.				
KW	transferase; Selenium; Selenocysteine; ATP-binding.				
FT	ACT_SITE 63 63				
FT	SE_CYS 66 66				
FT	SITE 66 66				
FT	IMPORTANT FOR CATALYTIC ACTIVITY (BY				

```

FT      NP_BIND      322      328      SIMILARITY).
FT      DOMAIN      2      9      ATP (POTENTIAL).
FT      DOMAIN      433      440      POLY-ALA.
SO      SEQUENCE      452 AA; 47786 MW; 3DA6F7250CFE80E4 CRC64;

Query Match      93.5%; Score 29; DB 1; Length 452;
Best Local Similarity 66.7%; Pred. No. 79;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 FXPXMR 6
DB      48 FSPSWR 53

RESULT 2
Y264_SYNY3
ID      Y264_SYNY3      STANDARD:      PRT: 477 AA.
AC      P73436;
DT      15-JUL-1998 (Rel. 36, Created)
DT      15-JUL-1998 (Rel. 36, Last sequence update)
DT      20-AUG-2001 (Rel. 40, Last annotation update)
DE      HYDROTHERICAL 54.0 KDA PROTEIN SLL1464.
CN      SLL1464.
OS      Synecocystis sp. (strain PCC 6803).
OC      Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
OX      NCBI_TaxID=1148;
RN      11
RP      SEQUENCE FROM N. A.
RX      MEDLINE=97061201; Pubmed=8905231;
RA      Kaneko T., Sato S., Kocani H., Tanaka A., Asamizu E., Nakamura Y.,
RA      Miyajima N., Hikosawa M., Sugita M., Sasamoto S., Kimura T.,
RA      Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA      Okumura S., Shimo S., Takeuchi C., Wada T., Watanabe A.,
RA      Yamada M., Yasuda M., Tabata S.;
RT      *Sequence analysis of the genome of the unicellular cyanobacterium
RT      Synecocystis sp. strain PCC6803. II. Sequence determination of the
RT      entire genome and assignment of potential protein-coding regions.*;
RL      DNA Res. 3:109-136(1996).
CC      -1- SIMILARITY: BELONGS TO THE UPF0061 FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: D90906; BAA17476.1;
CC      DR      InterPro: IPR003846; UPF0061.
CC      DR      Pfam: PF02696; UPF0061.1.
CC      KM      Hypothetical protein: Complete proteome.
CC      SO      SEQUENCE 477 AA; 54041 MW; B1F6899B1A6D613C CRC64;

Query Match      93.5%; Score 29; DB 1; Length 477;
Best Local Similarity 66.7%; Pred. No. 83;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 FXPXMR 6
DB      388 FSPSWR 393

RESULT 3
V70K_PLRV1
ID      V70K_PLRV1      STANDARD:      PRT: 639 AA.
AC      P17519;
DT      01-AUG-1990 (Rel. 15, Created)
DT      01-AUG-1990 (Rel. 15, Last sequence update)
DT      15-DEC-1998 (Rel. 37, Last annotation update)

```

```

DE      69.7 KDA PROTEIN (ORF 2).
OS      Potato leafroll virus (strain 1) (PLRV).
OC      Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
OC      Polerovirus.
OX      NCBI_TaxID=12046;
RN      11
RP      SEQUENCE FROM N.A.
RX      MEDLINE=89279282; Pubmed=2732710;
RA      Mayo M.A., Robinson D.J., Jolly C.A., Hyman L.;
RT      *Nucleotide sequence of potato leafroll luteovirus RNA.*;
RL      J. Gen. Virol. 70:1037-1051(1989).
CC      -1- SIMILARITY: 54% SIMILARITY TO BEET WESTERN YELLOW VIRUS ORF2.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: D00530; BAA00417.1;
CC      DR      EMBL: X14600; CAA32741.1;
CC      DR      PIR: JA0118; MWV070.
CC      DR      PIR: S24591; S24591.
CC      DR      InterPro: IPR000382; Luteo_ORF2.
CC      IR      Pfam: PF02122; Luteo_ORF2.1.
CC      IR      PRINTS: PR00913; LVIRUSORF2.
CC      SO      SEQUENCE 639 AA; 69629 MW; AFCEFB393BEE097 CRC64;

Query Match      93.5%; Score 29; DB 1; Length 639;
Best Local Similarity 66.7%; Pred. No. 11e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 FXPXMR 6
DB      615 FSPSWR 620

RESULT 4
HSPD_BRAJA
ID      HSPD_BRAJA      STANDARD:      PRT: 151 AA.
AC      O69241;
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      SMALL HEAT SHOCK PROTEIN HSPD.
CN      HSPD.
OS      Bradyrhizobium japonicum.
OC      Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC      Bradyrhizobium group; Bradyrhizobium.
OX      NCBI_TaxID=375;
RN      11
RP      SEQUENCE FROM N.A.
RA      Narberhaus F., Weiglhofer W., Fischer H.M., Hennecke H.;
RT      *Identification of the Bradyrhizobium japonicum degp gene as part of
RT      an operon containing small heat shock protein genes.*;
RL      Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.
CC      -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
CC      FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: AJ003064; CAA05835.1;
CC      DR      InterPro: IPR002068; Crystallin_HSP20.
CC      DR      Pfam: PF00011; HSP20; 1.

```

DR PROSITE: PS01031; HSP20; 1.
 KW Heat shock; Multigene family.
 SQ SEQUENCE 151 AA; 17272 MW; FB44EF94FB599EE4 CRC64;

Query Match 90.3%; Score 28; DB 1; Length 151;
 Best Local Similarity 66.7%; Pred. No. 45;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPXMR 6
 I I I I
 Db 6 FSPLMR 11

RESULT 5
 HSPB_BRAJA STANDARD; PRT; 151 AA.
 AC OR6110;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE SMALL HEAT SHOCK PROTEIN HSPH.
 GN HSPH.
 OS Bradyrhizobium japonicum.
 CC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 CC Bradyrhizobium group; Bradyrhizobium.
 CC NCBI_TaxID=375;
 OX NCBI_TaxID=375;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Muenchbach M., Nocker A., Narberhaus F.;
 RT "Occurrence of a superfamily of small heat shock proteins in
 RT Bradyrhizobium japonicum and other Rhizobium species: a plant-like
 RT phenomenon.";
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
 CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: AJ010144; CAA09014.1;
 DR InterPro: IPR002068; CrystalIn_HSP20.
 DR Pfam: PF00011; HSP20; 1.
 DR PROSITE: PS01031; HSP20; 1.
 KW Heat shock; Multigene family.
 SQ SEQUENCE 151 AA; 17098 MW; DFE9809310AF6A28 CRC64;

Query Match 90.3%; Score 28; DB 1; Length 151;
 Best Local Similarity 66.7%; Pred. No. 45;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPXMR 6
 I I I I
 Db 6 FSPLMR 11

RESULT 6
 HSPB_BRAJA STANDARD; PRT; 153 AA.
 AC P70918;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE SMALL HEAT SHOCK PROTEIN HSPB.
 GN HSPB.
 OS Bradyrhizobium japonicum.
 CC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

CC Bradyrhizobium group; Bradyrhizobium.
 CC NCBI_TaxID=375;
 OX NCBI_TaxID=375;
 RN [1]
 RP SEQUENCE FROM N.A.

KX MEDLINE-96404779; Pubmed-8808920;
 RA Narberhaus F., Weighthofer W., Fischer H.M., Hennecke H.;
 RT "The Bradyrhizobium japonicum rpoH gene encoding a sigma 32-like
 RT protein is part of a unique heat shock gene cluster together with
 RT groESL and three small heat shock genes.";
 RL J. Bacteriol. 178:5337-5346(1996).
 CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
 CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: U55047; AAC44756.1;
 DR InterPro: IPR002068; CrystalIn_HSP20.
 DR Pfam: PF00011; HSP20; 1.
 DR PROSITE: PS01031; HSP20; 1.
 KW Heat shock; Multigene family.
 SQ SEQUENCE 153 AA; 17170 MW; E2E7E248A69BFD38 CRC64;

Query Match 90.3%; Score 28; DB 1; Length 151;
 Best Local Similarity 66.7%; Pred. No. 45;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPXMR 6
 I I I I
 Db 7 FAPLMR 12

RESULT 7

VENV_BEV STANDARD; PRT; 233 AA.

AC P27904;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-AUG-1992 (Rel. 23, Last annotation update)
 DE ENVELOPE PROTEIN.

GN E.
 OS Berne virus (BEV).

CC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
 CC Coronaviridae; Torovirus.

CC NCBI_TaxID=11156;

OX NCBI_TaxID=11156;

RN [1]
 RP SEQUENCE FROM N.A.

KX MEDLINE-91220715; Pubmed-2024492;
 RA den Boon J.A., Snijder E.J., Locker J.K., Horzinek M.C.,
 RA Rottier P.J.M.;

RT "Another triple-spanning envelope protein among intracellularly
 RT budding RNA viruses: the torovirus E protein.";
 RT Virology 182:655-663(1991).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: X52505; CAA36747.1;
 DR PIR: A39869; VMWJBV.
 DR PIR: S15570; S15570.
 KW Envelope protein.

SQ SEQUENCE 233 AA: 26548 MW: A07A34DC539104BE CRC64:

Query Match 90.3%: Score 28: DB 1: Length 233:

Best Local Similarity 66.7%: Pred. No. 67:

Matches 4: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

OY 1 FXPYR 6

DB 60 FSPYR 65

RESULT 8

DPSD_CAEEL STANDARD: PRT: 333 AA.

AC 019949:

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE PUTATIVE PHOSPHATIDYL-SERINE DECARBOXYLASE PROENZYME (EC 4.1.1.65).

GN B0361.5.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodermidae; Caenorhabditis.

OX NCBI_TaxID=6239.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Du Z.;

RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.

CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYL-L-SERINE +

CC PHOSPHATIDYLETHANOLAMINE + CO(2).

CC -1- COFACTOR: REQUIRES A PYRUVYL GROUP FOR ITS ACTIVITY

CC (BY SIMILARITY).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: U00031: AKI8866.1: -

DR WormPep: B0361.5: CE00834.

DR InterPro: IPR003817: PS_Dcarboxylase.

DR Pfam: PF02666: PS_Dcarboxylase: 1.

KM Hypothetical protein: Phospholipid biosynthesis: Lyase: Decarboxylase;

KM Pyruvate: zymogen.

FT CHAIN 1 299

FT CHAIN (BY SIMILARITY).

FT CHAIN (BY SIMILARITY).

FT CHAIN (BY SIMILARITY).

FT SITE 299 300

FT SITE (BY SIMILARITY).

FT MOD_RES 300 300

FT MOD_RES (BY SIMILARITY).

FT SEQUENCE 333 AA: 37596 MW: 18CF04F2F31E1F34 CRC64:

Query Match 90.3%: Score 28: DB 1: Length 333:

Best Local Similarity 66.7%: Pred. No. 93:

Matches 4: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

OY 1 FXPYR 6

DB 36 FTPWR 41

RESULT 9

GLPK_MYCLE STANDARD: PRT: 508 AA.

AC 09CDB1:

UT 20-AUG-2001 (Rel. 40, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE GLYCEROL KINASE (EC 2.7.1.30) (ATP:GLYCEROL 3-PHOSPHOTRANSFERASE)

DE (GLYCEROKINASE) (CK).

GN GLPK OR ML2314.

OS Mycobacterium lepreae.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxID=1769.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-TN:

RX MEDLINE-21128732: PubMed-11234002;

RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,

RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,

RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,

RA Davies R.M., Devlin K., Duthoy S., Feltham T., Fraser A., Hamlin N.,

RA Holroyd S., Hornsby T., Jagers K., Lacroix C., Maclean J., Moule S.,

RA Murphy L., Oliver K., Quail M.A., Rajadram M.A., Rutherford K.M.,

RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,

RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,

RA Bartell B.G.;

RT "Massive gene decay in the leprosy bacillus";

RL Nature 409:1007-1011(2001).

CC -1- FUNCTION: KEY ENZYME IN THE REGULATION OF GLYCEROL UPTAKE AND

CC METABOLISM.

CC -1- CATALYTIC ACTIVITY: ATP + GLYCEROL = ADP + GLYCEROL 3-PHOSPHATE.

CC -1- PATHWAY: RATE-LIMITING STEP IN GLYCEROL UTILIZATION.

CC -1- SIMILARITY: BELONGS TO THE FUCOKINASE / GLUCOKINASE /

CC GLYCEROKINASE / XYLUOKINASE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: AL583925: CAC31830.1: -

DR Leproma: ML2314.

DR InterPro: IPR005577: FGGY_Kin.

DR Pfam: PF00370: FGGY_1.

DR PROSITE: PS00933: FGGY_KINASES_1.

DR PROSITE: PS00445: FGGY_KINASES_2.

KM Glycitol metabolism: transferase; Kinase; ATP-binding;

KM Complete proteome.

FT NP_BIND 157 169

FT NP_BIND (ATP PROBABLE).

FT SEQUENCE 508 AA: 54658 MW: A15379793ECF4039 CRC64:

Query Match 90.3%: Score 28: DB 1: Length 508:

Best Local Similarity 66.7%: Pred. No. 14+02:

Matches 4: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

OY 1 FXPYR 6

DB 360 FSPYR 365

RESULT 10

GLPK_MYCTU STANDARD: PRT: 517 AA.

AC 069664:

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE GLYCEROL KINASE (EC 2.7.1.30) (ATP:GLYCEROL 3-PHOSPHOTRANSFERASE)

DE (GLYCEROKINASE) (CK).

GN GLPK OR RV3696C OR MT3798 OR MTV025.044C.

OS Mycobacterium tuberculosis.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
 Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
 Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 Oliver K., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 Rutter S., Seeger K., Skellon J., Squares R., Squares S.,
 Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of mycobacterium tuberculosis from the
 complete genome sequence."
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 Kolony J.F., Nelson W.C., Umayan L.A., Ermolaeva M.D., Salzberg S.L.,
 Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 Bishal W.;
 RT "Whole genome comparison of mycobacterium tuberculosis clinical and
 laboratory strains."
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: KEY ENZYME IN THE REGULATION OF GLYCEROL UPTAKE AND
 METABOLISM.
 CC -1- CATALYTIC ACTIVITY: ATP + GLYCEROL = ADP + GLYCEROL 3-PHOSPHATE.
 CC -1- PATHWAY: RATE-LIMITING STEP IN GLYCEROL UTILIZATION.
 CC -1- SIMILARITY: BELONGS TO THE PUCOKINASE / GLUCOKINASE /
 GLYCEROKINASE / XYLOKINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AL022121; CAI18018.1; -
 CC EMBL: AF007177; AAK48165.1; -
 CC HSSP: P08859; IGLB.
 CC TIGR: MT3798; -
 DR TubercuList: RV3696C; -
 DR InterPro: IPR000577; FCGY_Kin.
 DR Pfam: PF00370; FCGY_1.
 DR PROSITE: PS00445; FCGY_KINASES_2; 1.
 DR PROSITE: PS00933; FCGY_KINASES_1; 1.
 KW Glyceral metabolism; Transferrase; Kinase; ATP-binding;
 KW Complete proteome.
 FT NP_BIND 165 177 ATP (PROBABLE).
 SQ SEQUENCE 517 AA; 55859 MW; 7E3F30DCF2A1E63E CRC64;

Query Match 90.3%; Score 28; DB 1; Length 517;
 Best Local Similarity 66.7%; Pred. No. 1.4e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EXPXMR 6
 DB 369 FAPYWR 374
 RESULT 11
 V70K_PLRVW STANDARD: PRT: 639 AA.
 AC P11622;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)

DT 01-AUG-1990 (Rel. 15, Last annotation update)
 DE 69.7 KDA PROTEIN (ORF 2).
 CS Potato leafroll virus (strain Wageningen) (PLRV).
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
 CC Poliovirus.
 OX NCBI_TaxID=12048;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=89171329; PubMed=2466700;
 RA van der Wilk F., Hutman M.J., Cornelissen B.J.C., Huttinga H.,
 KA Goldbach R.W.;
 RT "Nucleotide sequence and organization of potato leafroll virus
 genomic RNA."
 RL FEBS Lett. 245:51-56(1989).
 CC -1- SIMILARITY: 54% SIMILARITY TO BEET WESTERN YELLOWS VIRUS ORF2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: Y07496; CA68795.1; -
 DR PIR: S03547; S03547.
 DR InterPro: IPR000382; Luteo_ORF2.
 DR Pfam: PF0122; Luteo_ORF2_1.
 DR PRINTS: PR00913; LVIRUSORF2.
 SQ SEQUENCE 639 AA; 69676 MW; 92E1473FE3FEF148 CRC64;

Query Match 90.3%; Score 28; DB 1; Length 639;
 Best Local Similarity 66.7%; Pred. No. 1.7e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EXPXMR 6
 DB 615 FIPSWR 620

RESULT 12
 YAF7_CAEEL STANDARD: PRT: 199 AA.
 AC P52883;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 NE HYPOTHETICAL 23.2 KDA PROTEIN P46C5.7 IN CHROMOSOME II.
 GN F46C5.7.
 CS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 CC Rhabditidae; Peleoderinae; Caenorhabditis.
 CC NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Thomas K.;
 RT Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: 254281; CAA91049.1; -
 DR Wormpep: P46C5.7; CE03348.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 35 55 POTENTIAL.

FT TRANSMEM 57 77 POTENTIAL.
 FT TRANSMEM 94 114 POTENTIAL.
 FT TRANSMEM 131 151 POTENTIAL.
 SQ SEQUENCE 199 AA: 23247 MW: AC09222A692E3638 CRC64:

Query Match 87.1%; Score 27; DB 1; Length 199;
 Best Local Similarity 66.7%; Pred. No. 90;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 DB 22 FMPFMR 27

RESULT 13
 HO_FUGRU STANDARD; PRT: 277 AA.
 AC 073688;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE HEME OXYGENASE (EC 1.14.99.3) (HO).
 OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontidae; Takifugu.
 NCBI_TaxID=31033;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98163746; PubMed=9503016;
 RA Gottgens B., Gilbert J.G.R., Barton L.M., Aparicio S., Hawker K.,
 RA Mistry S., Vaudin M., King A., Bentley D., Elgar G., Green A.R.;
 RT "The pufferfish SLP-1 gene, a new member of the SCL/TAL-1 family of
 transcription factors".
 RL Genomics 48:52-62(1998).
 CC -1- FUNCTION: HEME OXYGENASE CLEAVES THE HEME RING AT THE ALPHA
 CC METHENE BRIDGE TO FORM BILIVERDIN. BILIVERDIN IS SUBSEQUENTLY
 CC CONVERTED TO BILIRUBIN BY BILIVERDIN REDUCTASE. UNDER
 CC PHYSIOLOGICAL CONDITIONS, THE ACTIVITY OF HEME OXYGENASE IS
 CC HIGHEST IN THE SPLEEN, WHERE SENESCENT ERYTHROCYTES ARE
 CC SEQUESTERED AND DESTROYED (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: HEME + 3 AH(2) + O(2) = BILIVERDIN + Fe(2+) +
 CC CO + 3 A + 3 H(2)O.
 CC -1- SUBCELLULAR LOCATION: MICROSOMAL (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE HEME OXYGENASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF022814; AAC41263.1;
 DR InterPro: IPR002051; Heme_oxygenase.
 DR Pfam: PF01126; Heme_oxygenase.1.
 DR PRINTS: PR00088; HAEMOXYGNASE.
 DR PROSITE: PS00593; HEME_OXYGENASE.1.
 KW Heme; Oxidoreductase; Microsome.
 FT BINDING 29 29 PROXIMAL HEME LIGAND (BY SIMILARITY).
 FT SEQUENCE 277 AA: 31211 MW: 77B3584699963F77 CRC64:

Query Match 87.1%; Score 27; DB 1; Length 277;
 Best Local Similarity 66.7%; Pred. No. 1.2e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 DB 101 FCPDWR 106

RESULT 14
 BCS1_YEAST STANDARD; PRT: 456 AA.

AC P32839; Q06404;
 DT 01-OCT-1993 (Rel. 37, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE BCS1 PROTEIN.
 GN BCS1 OR YDR375C OR D9481.17.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 NCBI_TaxID=4932;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9310976; PubMed=1327750;
 RA Nobrega F.G., Nobrega M.P., Tzagoloff A.;
 RT "BCS1, a novel gene required for the expression of functional Rieske
 RT iron-sulfur protein in Saccharomyces cerevisiae".
 RL EMBO J. 11:3821-3829(1992).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
 RA Favellio A., Fulton L., Gattung S., Greco T., Kirsten J.,
 RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
 RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
 RA Mardis E., Meneses S., Miller N., Nhan M., Pauley A., Peluso D.,
 RA Rifken L., Riles L., Tatch A., Trevasakis E., Vignati D.,
 RA Wilcox L., Wolman P., Vaudin M., Wilson R., Waterston R.;
 RT Submitted (JUN-1993) to the EMBL/GenBank/DBJ databases.
 RL (3)
 RP TOPOLOGY.
 RX MEDLINE=96174808; PubMed=8599931;
 RA Foelsch H., Guiard B., Neupert W., Stuart R.A.;
 RT "Internal targeting signal of the BCS1 protein: a novel mechanism of
 RT import into mitochondria".
 RL EMBO J. 15:479-487(1996).
 CC -1- FUNCTION: ESSENTIAL FOR THE EXPRESSION OF THE RIESKE IRON-SULFUR
 CC PROTEIN.
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND, MITOCHONDRIAL.
 CC -1- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: S47190; AAC09007.1;
 DR EMBL: U28373; AAB64811.1;
 DR PIR: S28411; S28411.
 DR SGD: S0002783; BCS1.
 DR InterPro: IPR003593; AAA.
 DR InInterPro: IPR001939; AAA_subfam.
 DR Pfam: PF00004; AAA.1.
 DR SMART: SM00382; AAA.1.
 DR PROSITE: PS00674; AAA.1.
 KW ATP-binding; Mitochondrion; Transmembrane.
 FT DOMAIN 1 44 MITOCHONDRIAL INTERMEMBRANE (PROBABLE).
 FT TRANSMEM 45 68 PROBABLE.
 FT DOMAIN 69 456 MITOCHONDRIAL MATRIX (PROBABLE).
 FT SITE 69 83 ATP (POTENTIAL).
 FT NP_BIND 267 274 ATP (POTENTIAL).
 FT CONFLICT 119 119 N -> F (IN REF. 1).
 FT CONFLICT 267 267 G -> S (IN REF. 1).
 FT CONFLICT 349 349 L -> Q (IN REF. 1).
 FT CONFLICT 351 351 G -> S (IN REF. 1).
 FT SEQUENCE 456 AA: 51107 MW: 17750D26B664ED04 CRC64:

Query Match 87.1%; Score 27; DB 1; Length 456;
 Best Local Similarity 66.7%; Pred. No. 1.9e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 EXPXMR 6
 1 1 1 1
 DB 209 FCPXMR 214

RESULT 15

RBL2_RHOSH STANDARD; PRT; 459 AA.
 AC P29278;
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN 2 (EC 4.1.1.39)
 DE (RUBISCO LARGE SUBUNIT).
 GN CBRL2 OR RBP.
 OS Rhodospirillum rubrum (Rhodospirillum rubrum).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum group;
 OC Rhodospirillum.
 OX NCBI_TaxID=1063;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wagner S.J., Stevens S.E. Jr., Nixon B.T., Lambert D.H.,
 RA Olveij R.G. Jr., Tabita F.R.;
 RT Nucleotide and deduced amino acid sequence of the Rhodospirillum
 RT rubrum gene encoding form II ribulose-1,5-bisphosphate
 RT carboxylase/oxygenase and comparison with other deduced forms I and II
 RT sequences.";
 RT FEMS Microbiol. Lett. 55:217-222(1988).
 RL [2]
 RP SEQUENCE OF 1-36 FROM N.A.
 RX MEDLINE-92041881; PubMed-1939098;
 RA Chen J.-H., Gibson J.L., McCue L.A., Tabita F.R.;
 RT Identification, expression, and deduced primary structure of
 RT transketolase and other enzymes encoded within the form II CO2
 RT fixation operon of Rhodospirillum rubrum.";
 RL J. Biol. Chem. 266:20447-20452(1991).
 CC -1- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
 CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
 CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
 CC ACTIVE SITE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) =
 CC 2,3-PHOSPHO-D-GLYCERATE.
 CC -1- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) =
 CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
 CC -1- SUBUNIT: IN CONTRAST TO FORM I GENES, THE FORM II RUBISCO ARE
 CC COMPOSED OF ONLY LARGE SUBUNITS.
 CC -1- MISCELLANEOUS: THIS PROTEIN IS ENCODED WITHIN THE FORM II
 CC RIBULOSE-BISPHOSPHATE CARBOXYLASE OPERON.
 CC -1- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M68914; AAA26158.1; ALU_SEQ.
 CC PIR: E41080; E41080.
 CC HSP: P04718; IRBA.
 CC InterPro: IPR000685; Rubisco_large.
 CC Pfam: PF00016; Rubisco_large; 1.
 CC PROSITE: PS00157; Rubisco_LARGE; 1.
 CC Photosynthesis; Carbon dioxide fixation; Photorespiration;

FW Lyase; Oxidoreductase; Monooxygenase; Multigene family.
 FT ACT_SITE 191 191 BINDING OF CO(2) ACTIVATES THE ENZYME.
 SQ SEQUENCE 459 AA; 50519 MW; 299ABAA836BD683E CRC64;

Query Match 87.1%; Score 27; DB 1; Length 459;
 Best Local Similarity 66.7%; Pred. No. 1.9e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 EXPXMR 6
 1 1 1 1
 DB 447 FPGMR 452

Search completed: February 27, 2002, 11:42:49
 Job time: 548 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using SW model

Run on: February 27, 2002, 11:50:14 : Search time 281.76 seconds
(without alignments)
3.115 Million cell updates/sec

Title: US-09-446-109a-18
Perfect score: 31
Sequence: 1 FXPXWR 6

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database:

SPTREMBL_17:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_mammal:*
- 6: sp_invertebrate:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_protent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	93.5	296	2	033353 mycobacteri
2	29	93.5	354	3	060121 schizosacch
3	29	93.5	563	2	P72421
4	29	93.5	572	12	093866
5	28	90.3	135	2	09PB73
6	28	90.3	141	2	09WVK5
7	28	90.3	143	5	096402
8	28	90.3	160	4	095530
9	28	90.3	160	4	098029
10	28	90.3	211	2	033867
11	28	90.3	218	4	095726
12	28	90.3	221	10	09XJ12
13	28	90.3	236	2	005115
14	28	90.3	241	5	09VLD3
15	28	90.3	250	2	09EUM2
16	28	90.3	254	2	049185
17	28	90.3	254	2	09F1E6
18	28	90.3	264	2	09FDH9
19	28	90.3	272	11	09DA91

20	28	90.3	297	10	09ZOC7	09ZQC,
21	28	90.3	311	5	044789	044789 cae.
22	28	90.3	317	2	09L4E0	09L4E0 bacillu
23	28	90.3	325	2	09L4E9	09L4E9 bacillus c
24	28	90.3	337	1	09HP24	09HP24 halobacteri
25	28	90.3	357	5	026862	026862 trypanosom
26	28	90.3	389	2	0910K8	0910K8 pseudomonas
27	28	90.3	407	2	09AA59	09AA59 caulobacter
28	28	90.3	410	10	041018	041018 plus strob
29	28	90.3	475	1	026479	026479 methanobact
30	28	90.3	483	2	09A8C6	09A8C6 caulobacter
31	28	90.3	488	10	041017	041017 plus strob
32	28	90.3	490	2	09A3J3	09A3J3 caulobacter
33	28	90.3	507	2	09RJM2	09RJM2 streptomyce
34	28	90.3	512	2	09ADA7	09ADA7 streptomyce
35	28	90.3	542	10	09SHX0	09SHX0 arabidopsi
36	28	90.3	552	2	006335	006335 mycobacteri
37	28	90.3	559	4	09H5P1	09H5P1 homo sapien
38	28	90.3	624	4	09C0D7	09C0D7 homo sapien
39	28	90.3	633	4	09Y4W3	09Y4W3 homo sapien
40	28	90.3	635	2	0927B3	0927B3 chlamydia p
41	28	90.3	635	2	09JSA3	09JSA3 chlamydia p
42	28	90.3	639	12	084836	084836 potato leaf
43	28	90.3	703	10	09M1P7	09M1P7 arabidopsi
44	28	90.3	781	4	095560	095560 homo sapien
45	28	90.3	791	4	09C0B6	09C0B6 homo sapien

ALIGNMENTS

RESULT 1

C33353 ID 033353 PREDIMINARY: PRT: 296 AA.

BT 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)

DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)

FE HYPOTHEtical 32.0 KDA PROTEIN (ORF2).

OS Mycobacterium tuberculosis.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomyetales; Corynebacteriineae; Mycobacteriaceae; Mycobacterium.

OX NCBI_TaxID=1773;

RN [1]

RP SEQUENCE FROM N.A.

RC TRANSPOSOM-1S1547;

KX MEDLINE=99121044; PubMed=9922269;

KA Fang Z., Doig C., Morrison N., Walt B., Forbes K.J.;

FT "Characterization of IS1547, a new member of the IS900 family in the

RT Mycobacterium tuberculosis complex, and its association with IS6110."

PL J. Bacteriol. 181:1021-1024(1999).

EN [2]

RP SEQUENCE FROM N.A.

RC TRANSPOSOM-1S1547;

KX MEDLINE=97157336; PubMed=9003621;

KA Fang Z., Forbes K.J.;

FT "A Mycobacterium tuberculosis IS6110 preferential locus (ipl) for

RT insertion into the genome."

PL J. Clin. Microbiol. 35:479-481(1997).

DR EMBL: Y13470; CAA73869.1;

KW Transposable element; Hypothetical protein.

SO SEQUENCE 296 AA; 31996 MW; 3CD4CHC956653F0 CRC64;

Query Match 93.5%; Score 29; DB 2; Length 296;
Best Local Similarity 66.7%; Pred. No. 3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXWR 6
DB 263 FSPXWR 268

RESULT 2
 060121 PRELIMINARY: PRT: 354 AA.
 AC 060121:
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE HYPOPHETICAL 39.3 KDA PROTEIN C1605.07C IN CHROMOSOME II.
 GN SPBC1605.07C.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 NCBI_TaxID=4896;
 RX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Lyne M., Wood V., Rajandream M.A., Barrell B.G., Beck A.,
 Reinhardt R.,
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE BAND 7 / MEC-2 FAMILY.
 DR EMBL; AL023554; CAA19027.1; -
 DR InterPro: IPR001107; Band_7.
 DR InterPro: IPR001972; Stomatin.
 DR Pfam: PF01145; Band_7; 1.
 DR SMART: SM00244; PHE; 1.
 DR PROSITE: PS01270; BAND_7; FALSE_NEG.
 KW Hypothetical protein.
 SQ SEQUENCE 354 AA: 39274 MW: 6680966A12BDC030 CRC64:

Query Match 93.5%: Score 29; DB 3; Length 354;
 Best Local Similarity 66.7%: Pred. No. 3.5e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 1 1 1 1
 DB 40 FXPXMR 45

RESULT 3
 P72421 PRELIMINARY: PRT: 563 AA.
 AC P72421:
 DT 01-FEB-1997 (TREMBLrel. 02, Created)
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE (SEROTYPE TYPHIMURIUM).
 GN MKFB.
 OS Salmonella typhimurium.
 OC Plasmid.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 NCBI_TaxID=602;
 RX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SEROTYPE TYPHIMURIUM, C5;
 RX MEDLINE=90161559; Pubmed=2696057;
 RA Norel F., Pisano M.R., Nicoli J., Popoff M.Y.;
 RT "Nucleotide sequence of the plasmid-borne virulence gene mktB from
 RT Salmonella typhimurium."
 RL Res. Microbiol. 140:455-457(1989).
 GN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SEROTYPE TYPHIMURIUM, C5;
 RA Norel F.;
 DR Submitted (JAN-1992) to the EMBL/GenBank/DBJ databases.
 DR EMBL; X57096; CAA40380.1; -
 DR InterPro: IPR003284; Sal_SpV.
 DR PRINTS; PRO1341; SALSPVBPROT.
 KW Plasmid.
 SQ SEQUENCE 563 AA: 62586 MW: DA3F093AB82A1C0A CRC64:

Query Match 93.5%: Score 29; DB 2; Length 563;
 Best Local Similarity 66.7%: Pred. No. 5.5e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 1 1 1 1
 DB 50 FAPXMR 55

RESULT 4
 09386 PRELIMINARY: PRT: 572 AA.
 AC 09386:
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE ORF19 CHITINASE.
 OS Spodoptera exigua nucleopolyhedrovirus.
 CC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
 CC Nucleopolyhedrovirus.
 NCBI_TaxID=10454;
 RX NCBI_TaxID=10454;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=20036646; Pubmed=10567663;
 RX Lukel W.F., van Strien E.A., Hieldens J.G., Broer R., Zuidema D.,
 RA Goldbach R.W., Vlask J.M.;
 RT "Sequence and organization of the spodoptera exigua multicapsid
 RT nucleopolyhedrovirus genome."
 RL J. Gen. Virol. 80:3289-3304(1999).
 GN [2]
 RP SEQUENCE FROM N.A.
 RC Lukel W.F., van Strien E.A., Hieldens J.G.M., Broer R., Zuidema D.,
 RA Goldbach R.W., Vlask J.M.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF169823; AAF33549.1; -
 DR InterPro: IPR000886; ER_target.
 DR InterPro: IPR001223; Glyco_hydro_18.
 DR InterPro: IPR002173; PfkB.
 DR Pfam: PF00704; Glyco_hydro_18; 1.
 DR PROSITE: PS00014; ER_TARGET; UNKNOWN_1.
 DR PROSITE: PS00583; PFKB_KINASES_1; UNKNOWN_1.
 SQ SEQUENCE 572 AA: 63508 MW: A678AAD14021A014 CRC64:

Query Match 93.5%: Score 29; DB 12; Length 572;
 Best Local Similarity 66.7%: Pred. No. 5.5e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 1 1 1 1
 DB 401 FAPXMR 406

RESULT 5
 09PB73 PRELIMINARY: PRT: 135 AA.
 AC 09PB73:
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE HYPOPHETICAL PROTEIN XF2271.
 GN XF2271.
 GN Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xylella.
 NCBI_TaxID=2371;
 RX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9A5C;
 RX MEDLINE=20365717; Pubmed=10910347;
 RA Stimpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,

RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carier H.,
 RA Colauco N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
 RA Facchini A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,
 RA Gargiel M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hohlseil J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.P.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A.J., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Pasquero J.B.,
 RA Peixoto R.B., Pereira G.A.G., Pereira H.A., Jr., Pasquero J.B.,
 RA Quaglio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E., Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A.Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Zeldanis J., Setubal J.C.,
 RA "The genome sequence of the plant pathogen Xylella fastidiosa";
 RL Nature 406:151-159(2000).
 KW EMBL: AE004039; AAF85070.1; -
 RT Hypothetical protein: Complete proteome.
 SQ SEQUENCE 135 AA; 15203 MW; 99E7FE0EA22D05244 CRC64;

Query Match 90.3%; Score 28; DB 2; Length 135;
 Best Local Similarity 66.7%; Pred. No. 2.2e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 I I I I
 Db 79 FVPMWR 84

RESULT 6
 O9MWK5 PRELIMINARY; PRT; 141 AA.
 AC O9MWK5;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE HYPOHETICAL 16.0 KDA PROTEIN.
 OS Synechococcus sp. (strain PCC 7002) (Agmenellum quadruplicatum).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
 OX NCBI_TaxID=32049;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PCC7002, PR6000;
 RA Schlink K., Kowalik K.V., Bryant D.A.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF136577; AAD26588.1; -
 DR HSSP: P00210; 1FXR.
 DR InterPro: IPR001080; 3FE4S_ferrdxn.
 DR PRINTS: PR00352; 3FE4SFRDOXIN.
 KW Hypothetical protein.
 SQ SEQUENCE 141 AA; 16034 MW; 85872F258B76FCE2 CRC64;

Query Match 90.3%; Score 28; DB 2; Length 141;
 Best Local Similarity 66.7%; Pred. No. 2.3e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 I I I I
 Db 122 FSPGWR 127

RESULT 7
 O96402 PRELIMINARY; PRT; 143 AA.
 AC O96402;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
 DE ERYTHROCYTE MEMBRANE PROTEIN 1 (FRAGMENT).
 GN FCR3L2VAR6.
 OS Plasmodium falciparum.
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Chen Q., Wahlgren M.;
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF039278; AAD02161.1; -
 FT NON_TER 1 1
 FT NON_TER 1 1
 SQ SEQUENCE 143 AA; 16818 MW; A8641F68BE09B3CE CRC64;

Query Match 90.3%; Score 28; DB 5; Length 143;
 Best Local Similarity 66.7%; Pred. No. 2.4e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 I I I I
 Db 2 FNPTWR 7

RESULT 8
 O95530 PRELIMINARY; PRT; 160 AA.
 AC O95530;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
 DE DJ281I8.2 (PUTATIVE NOVEL PROTEIN SIMILAR TO KIAA0323 AND WORM
 IE C30F12.1) (FRAGMENT).
 GN DJ281I8.2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Masiregini-Mohammadi M.;
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL031133; CAA20017.1; -
 FT NON_TER 1 1
 FT NON_TER 1 1
 SQ SEQUENCE 160 AA; 18983 MW; EC9F622FCCFAEF3 CRC64;

Query Match 90.3%; Score 28; DB 4; Length 160;
 Best Local Similarity 66.7%; Pred. No. 2.6e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 I I I I
 Db 30 FVPMWR 35

RESULT 9
 O9BQZ9 PRELIMINARY; PRT; 160 AA.
 AC O9BQZ9;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DT DJ423B22.1 (NOVEL PROTEIN SIMILAR TO KIAA0323, KIAA0615 AND C. ELIACANS
 IE C30F12.1) (FRAGMENT).
 GN DJ423B22.1

GN D1423822.1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Pearce A.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL034379; CAC36083.1; -.
 FT NON_TER 1
 FT NON_TER 160
 SO SEQUENCE: 160 AA: 18953 MW: 538E2B52DAED6C0 CRC64:

Query Match 90.3%; Score 28; DB 4; Length 160;
 Best Local Similarity 66.7%; Pred. NO. 2.6e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 1 1 1 1
 DB 30 FVPSWR 35

RESULT 10
 ID 033867 PRELIMINARY; PRT; 211 AA.
 AC 033867;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE SIMILAR TO ANSR GENBANK ACCESSION NUMBER Q07683.
 OS Bacillus pumilus (Bacillus mesentericus).
 OC Plasmid PSH1452.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 NCBI_TaxID=1408;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SH1451;
 RA Thomas C.M., Hasnain S.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U53767; AAB71491.1; -.
 DR InterPro: IPR001387; HTH_3.
 DR Pfam: PF01381; HTH_3; 1.
 DR SMART: SM00530; HTH_XRE; 1.
 KW Plasmid.
 SO SEQUENCE 211 AA: 24620 MW: 49BD5954BD35E83A CRC64:

Query Match 90.3%; Score 28; DB 2; Length 211;
 Best Local Similarity 66.7%; Pred. NO. 3.4e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 1 1 1 1
 DB 199 F1PTWR 204

RESULT 11
 ID 095726 PRELIMINARY; PRT; 218 AA.
 AC 095726;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
 DE HYPOTHEICAL 25.5 KDA PROTEIN (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RC TISSUE=BRAIN.
 RA Park S.H., Kim C., Kang Y.H., Chung H.S., Kim H.;
 RT "Expression of novel gene mapped on chromosome 7q22."
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U67037; AAD09521.1; -.
 KW Hypothetical protein.
 FT NON_TER 1
 FT NON_TER 218
 SO SEQUENCE 218 AA: 25458 MW: C0896F5ED4C5627F CRC64:

Query Match 90.3%; Score 28; DB 4; Length 218;
 Best Local Similarity 66.7%; Pred. NO. 3.5e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 1 1 1 1
 DB 202 FDPXMR 207

RESULT 12
 ID 09XJ12 PRELIMINARY; PRT; 221 AA.
 AC 09XJ12;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE HYPOTHEICAL PROTEIN.
 OS Oryza sativa (Rice).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 CC Ehrhartoideae; Oryzaceae; Oryza.
 NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPONBARE;
 RA Sasaki T., Nagamura Y., Yamamoto K.;
 RT "Oryza sativa niponbare (GA3) genomic DNA, chromosome 6, PAC
 clone: P0680A03."
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB023482; BAA78741.1; -.
 SO SEQUENCE 221 AA: 24502 MW: DC24FF370660BC15 CRC64:

Query Match 90.3%; Score 28; DB 10; Length 221;
 Best Local Similarity 66.7%; Pred. NO. 3.5e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 FXPXMR 6
 1 1 1 1
 DB 203 FRPTWR 208

RESULT 13
 ID 005115 PRELIMINARY; PRT; 236 AA.
 AC 005115;
 DT 01-JUL-1997 (TREMBLrel. 04, Created)
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE PUTATIVE GLYCERATE KINASE AND PYRUVATE KINASE (PYKA) GENES, COMPLETE
 CDS (PYKA).
 OS Methylobacterium extorquens.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Methylobacterium group; Methylobacterium.
 OX NCBI_TaxID=408;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AM1;
 RA MEDLINE=97386438; Pubmed=9244287;
 RA Chistoserdova L., Lidstrom M.E.;
 RT "Identification and mutation of a gene required for glycerate kinase
 activity from a facultative methylotroph, Methylobacterium extorquens

RT AM1."
 RL J. Bacteriol. 179:4946-4948(1997).
 DR EMBL: U87316; AAB66495.1;
 SQ SEQUENCE 236 AA; 26139 MW; 7F48376B148B3676 CRC64;

Query Match 90.3%; Score 28; DB 2; Length 236;
 Best Local Similarity 66.7%; Pred. No. 3.7e+02;
 Matches 4: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EXPXMR 6
 1 1 1 1
 DB 135 FRPVMR 140

RESULT 14
 ID 09VLD3 PRELIMINARY; PRT: 241 AA.
 AC 09VLD3;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE CG3694 PROTEIN.
 CN CG3694.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NX NCBI_TaxId=7227;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Gelinkler S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatidis P.C., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Galbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaitli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacheb J.M.,
 RA Palazozo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Sidenkimos I., Simpson M., Skupski M.P., Smith T.,
 RA Spletter E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Wodtke T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 DR EMBL: AEO03624; AAF52759.1;
 DK Flybase: FBgn0032107; CG3694.

SQ SEQUENCE 241 AA; 27071 MW; 200925FE35C3F733 CRC64;

Query Match 90.3%; Score 28; DB 5; Length 241;
 Best Local Similarity 66.7%; Pred. No. 3.8e+02;
 Matches 4: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EXPXMR 6
 1 1 1 1
 DB 68 FRPVMR 73

RESULT 15
 ID 09EUM2 PRELIMINARY; PRT: 250 AA.
 AC 09EUM2;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE YCG4L.
 CN YCG4L.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Corynebacteriaceae;
 OC Corynebacterium.
 NX NCBI_TaxId=1718;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC31830;
 RA Tauch A., Puhler A., Kalinowski J.,
 RT "DNA sequence and genetic organization of the integron-carrying R-
 plasmid pCG4 of Corynebacterium glutamicum."
 DR Submitted (Jul-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF164956; AAC00294.1;
 DR InterPro: IPR002145; COPC_HTH_4.
 DR Pfam: PF01402; HTH_4; 1.
 RX Plasmid.
 SQ SEQUENCE 250 AA; 28804 MW; 195C86367B483157 CRC64;

Query Match 90.3%; Score 28; DB 2; Length 250;
 Best Local Similarity 66.7%; Pred. No. 4e+02;
 Matches 4: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 EXPXMR 6
 1 1 1 1
 DB 214 FRPVMR 219

Search completed: February 27, 2002, 11:50:14
 Job time: 993 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:36:08 ; Search time 132.19 Seconds
(Without alignments)
1.021 Million cell updates/sec

Title: US-09-446-109a-18
Perfect score: 31
Sequence: 1 FXPXWR 6

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A-COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B-COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A-COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B-COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PTUS-COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	Query Match	Length	DB ID	Description
1	28	90.3	527	3	US-08-907-229-2
2	28	90.3	1536	4	US-09-413-814-10
3	28	90.3	2232	4	US-09-091-219-25
4	28	90.3	2247	4	US-09-091-219-25
5	27	87.1	6	1	US-08-215-137-13
6	27	87.1	456	4	US-09-268-364-21
7	26	83.9	42	2	US-08-766-858A-27
8	26	83.9	254	4	US-09-362-831-3
9	26	83.9	273	2	US-08-997-080-75
10	26	83.9	273	2	US-08-997-362-75
11	26	83.9	273	3	US-08-873-970-75
12	26	83.9	273	4	US-09-095-855-75
13	26	83.9	370	2	US-08-997-080-194
14	26	83.9	370	2	US-08-997-362-194
15	26	83.9	370	4	US-09-095-855-194
16	26	83.9	410	4	US-09-238-480-2
17	26	83.9	604	2	US-08-735-041A-4
18	26	83.9	604	3	US-09-190-476B-4
19	26	83.9	604	3	US-09-190-889A-4
20	26	83.9	604	4	US-09-190-938A-4
21	26	83.9	701	2	US-08-735-041A-6
22	26	83.9	701	3	US-09-190-476B-6
23	26	83.9	701	3	US-09-190-889A-6
24	26	83.9	701	4	US-09-190-938B-6
25	25	80.6	38	1	US-08-176-500-118
26	25	80.6	38	1	US-08-471-052A-118
27	25	80.6	38	1	US-08-189-331-118

28	25	80.6	38	2	US-08-471-939-118	Sequence 118, App
29	25	80.6	38	2	US-08-471-800-118	Sequence 118, App
30	25	80.6	38	2	US-08-471-068-118	Sequence 118, App
31	25	80.6	175	4	US-09-060-726A-2	Sequence 2, Appl
32	25	80.6	631	1	US-08-487-890A-115	Sequence 115, App
33	25	80.6	631	2	US-08-478-435-115	Sequence 115, App
34	25	80.6	631	2	US-08-337-483-115	Sequence 115, App
35	25	80.6	631	2	US-08-478-373-115	Sequence 115, App
36	25	80.6	631	3	US-08-474-671-115	Sequence 115, App
37	25	80.6	631	3	US-08-483-577A-115	Sequence 115, App
38	25	80.6	631	4	US-08-897-438-115	Sequence 115, App
39	25	80.6	743	4	US-09-077-334B-2	Sequence 2, Appl
40	24	77.4	10	3	US-08-159-339A-448	Sequence 448, App
41	24	77.4	15	2	US-08-553-257A-52	Sequence 52, Appl
42	24	77.4	175	1	US-08-414-926A-8	Sequence 8, Appl
43	24	77.4	175	2	US-08-926-922B-8	Sequence 8, Appl
44	24	77.4	175	3	US-09-253-682-8	Sequence 8, Appl
45	24	77.4	175	4	US-09-527-657-8	Sequence 8, Appl

ALIGNMENTS

```

RESULT 1
US-08-907-229-2
: Sequence 2, Application US/08907229A
: Patent No. 6072048
: GENERAL INFORMATION:
: APPLICANT: Riley, Lee W.
: TITLE OF INVENTION: DNA MOLECULE ENCODING FOR CELLULAR UPTAKE OF
: FILE REFERENCE: 19603/1531
: CURRENT APPLICATION NUMBER: US/08/907,229A
: EARLIER FILING DATE: 1997-08-06
: EARLIER APPLICATION NUMBER: 60/040,097
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 2
: LENGTH: 527
: TYPE: PRT
: ORGANISM: Mycobacterium tuberculosis
: US-08-907-229-2

Query Match      90.3%  Score 28:  DB 3:  Length 527:
Best Local Similarity 66.7%:  Pred. No. 2.9e+02:
Matches 4:  Conservative 0:  Mismatches 2:  Indels 0:  Gaps 0:

Oy      1 FXPXWR
Db      3 FGPSWR 8

RESULT 2
US-09-413-814-10
: Sequence 10, Application US/09413814
: Patent No. 6225064
: GENERAL INFORMATION:
: APPLICANT: Gaselischschaff fuer Biotechnologische Forschung mbH
: APPLICANT: Bristol-Myers Squibb, CO.
: APPLICANT: Beyer, Stefan
: APPLICANT: Bloeker, Helmut
: APPLICANT: Brandt, Petra
: APPLICANT: Cino, Paul M
: APPLICANT: Dougherty, Brian A
: APPLICANT: Goldberg, Steven I.
: APPLICANT: Hoffe, Gerhard
: APPLICANT: Mueller, Joachim
: APPLICANT: Reichenbach, Hans
: TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
: FILE REFERENCE: PCT/US 99/23535

```

```

: CURRENT APPLICATION NUMBER: US/09/413,814
: CURRENT FILING DATE: 1999-10-07
: EARLIER APPLICATION NUMBER: DE 198 46 493.2
: EARLIER FILING DATE: 1998-10-09
: NUMBER OF SEQ ID NOS: 107
: SOFTWARE: Patentln Ver. 2.1
: SEQ ID NO 10
: LENGTH: 1536
: TYPE: PRT
: ORGANISM: Sorangium cellulosum
US-09-413-814-10

```

```

Query Match          90.3%; Score 28; DB 4; Length 1536;
Best Local Similarity 66.7%; Pred. No. 8.1e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 1 FXPXR 6
DB 896 FLPAWR 901

```

```

RESULT 3
US-09-091-219-25
: Sequence 25; Application US/09091219
: Patent No. 6171592
: GENERAL INFORMATION:
: APPLICANT: STUDDERT, Michael J.
: APPLICANT: CRAAB, Brendan S.
: APPLICANT: FENG, Li
: TITLE OF INVENTION: EQUINE RHINOVIRUS 1 PROTEINS
: FILE REFERENCE: 040268/0151
: CURRENT APPLICATION NUMBER: US/09/091,219
: CURRENT FILING DATE: 1998-10-05
: EARLIER APPLICATION NUMBER: PCT/AU96/00815
: EARLIER FILING DATE: 1996-12-18
: EARLIER APPLICATION NUMBER: AU PN7201
: EARLIER FILING DATE: 1995-12-18
: NUMBER OF SEQ ID NOS: 25
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 25
: LENGTH: 2232
: TYPE: PRT
: ORGANISM: equine rhinovirus 1
US-09-091-219-25

```

```

Query Match          90.3%; Score 28; DB 4; Length 2232;
Best Local Similarity 66.7%; Pred. No. 1.1e+03;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 1 FXPXR 6
DB 2227 FVPTWR 2232

```

```

RESULT 4
US-09-091-219-2
: Sequence 2; Application US/09091219
: Patent No. 6171592
: GENERAL INFORMATION:
: APPLICANT: STUDDERT, Michael J.
: APPLICANT: CRAAB, Brendan S.
: APPLICANT: FENG, Li
: TITLE OF INVENTION: EQUINE RHINOVIRUS 1 PROTEINS
: FILE REFERENCE: 040268/0151
: CURRENT APPLICATION NUMBER: US/09/091,219
: CURRENT FILING DATE: 1998-10-05
: EARLIER APPLICATION NUMBER: PCT/AU96/00815
: EARLIER FILING DATE: 1996-12-18
: EARLIER APPLICATION NUMBER: AU PN7201
: EARLIER FILING DATE: 1995-12-18
: NUMBER OF SEQ ID NOS: 25

```

```

: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 2
: LENGTH: 2247
: TYPE: PRT
: ORGANISM: equine rhinovirus 1
US-09-091-219-2

```

```

Query Match          90.3%; Score 28; DB 4; Length 2247;
Best Local Similarity 66.7%; Pred. No. 1.2e+03;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 1 FXPXR 6
DB 2227 FVPTWR 2232

```

```

RESULT 5
US-08-215-137-13
: Sequence 13; Application US/08215137
: Patent No. 5614370
: GENERAL INFORMATION:
: APPLICANT: Konleatis, Zenon
: APPLICANT: Siciliano, Salvatore J
: APPLICANT: Springer, Martin S
: TITLE OF INVENTION: ASSAY TO IDENTIFY HUMAN C5A ANTAGONISTS
: TITLE OF INVENTION: AND AGONISTS
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Merck & Co., Inc.
: STREET: 126 Lincoln Ave., P.O. Box 2000
: CITY: Rahway
: STATE: New Jersey
: COUNTRY: United States of America
: ZIP: 07065
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/215,137
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Bencen, Gerard H
: REGISTRATION NUMBER: 35,746
: TELECOMMUNICATION INFORMATION: 19108
: TELEPHONE: (908) 594-3901
: TELEFAX: (908) 594-4720
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: C-terminal
: FEATURE:
: NAME/key: Modified-site
: LOCATION: 1
: OTHER INFORMATION: /label- BH
: OTHER INFORMATION: /note- "either the natural phenylalanine amino
: OTHER INFORMATION: terminus of the Bolton-Hunter modified peptide
: OTHER INFORMATION: having the group 3-(p-hydroxyphenyl)propionyl group
: FEATURE:
: NAME/key: Modified-site
: LOCATION: 4
: OTHER INFORMATION: /label- dcha
: OTHER INFORMATION: /note- "D-cyclohexylalanine"

```


FEATURE:
NAME/KEY: Modified-site
LOCATION: 6
OTHER INFORMATION: /label= darg
OTHER INFORMATION: /note= "d-arginine"
US-08-215-137-13

Query Match 87.1%; Score 27; DB 1; Length 6;
Best Local Similarity 83.3%; Pred. No. 1.6e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FXPXWR 6
1 1111
DB 1 FXPXWR 6

RESULT 6
US-09-268-364-21
Sequence 21, Application US/09268364A
Patent No. 6204063
GENERAL INFORMATION:
APPLICANT: Allen, Steve
APPLICANT: Lee, Jian Ming
APPLICANT: Lightner, Jonathan
APPLICANT: Odell, Joan
TITLE OF INVENTION: PLANT GLYCOLYSIS AND RESPIRATION ENZYMES
FILE REFERENCE: BB-1154
CURRENT APPLICATION NUMBER: US/09/268-364A
CURRENT FILING DATE: 1999-03-15
EARLIER APPLICATION NUMBER: 60/079,387
EARLIER FILING DATE: March 16, 1998
NUMBER OF SEQ ID NOS: 21
SOFTWARE: Microsoft Office 97
SEQ ID NO 21
LENGTH: 456
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-268-364-21

Query Match 87.1%; Score 27; DB 4; Length 456;
Best Local Similarity 66.7%; Pred. No. 3.9e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXWR 6
1 1 11
DB 209 FGPEWR 214

RESULT 7
US-08-766-858A-27
Sequence 27, Application US/08766858A
Patent No. 5935782
GENERAL INFORMATION:
APPLICANT: Cianciocto, Nicholas P.
APPLICANT: Hickey, Erin K.
APPLICANT: O'Connell, William A.
TITLE OF INVENTION: METHODS AND MATERIALS FOR DETECTING
TITLE OF INVENTION: LECTIONELLA PNEUMOPHILA
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross P.C.
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/766-858A
FILING DATE: 13-DEC-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/011,545
FILING DATE: 13-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Crook, Wanneil M.
REGISTRATION NUMBER: 31,071
REFERENCE/DOCKET NUMBER: 3501-4-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 42 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-766-858A-27

Query Match 83.9%; Score 26; DB 2; Length 42;
Best Local Similarity 50.0%; Pred. No. .64;
Matches 3; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXWR 6
1 1 1:
DB 20 FSPXWK 25

RESULT 8
US-09-362-831-3
Sequence 3, Application US/09362831
Patent No. 6306400
GENERAL INFORMATION:
APPLICANT: BUBLLOT et al.
TITLE OF INVENTION: AVIAN RECOMBINANT LIVE VACCINE USING, AS VECTOR, THE
TITLE OF INVENTION: AVIAN INFECTIOUS LARYNGOTRACHEITIS VIRUS
FILE REFERENCE: 454313-2520
CURRENT APPLICATION NUMBER: US/09/362-831
CURRENT FILING DATE: 1999-07-28
NUMBER OF SEQ ID NOS: 28
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 254
TYPE: PRT
ORGANISM: Infectious laryngotracheitis virus
US-09-362-831-3

Query Match 83.9%; Score 26; DB 4; Length 254;
Best Local Similarity 66.7%; Pred. No. 3.5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXWR 6
1 1 11
DB 186 FGPPWR 191

RESULT 9
US-08-997-080-75
Sequence 75, Application US/08997080
Patent No. 5968524
GENERAL INFORMATION:
APPLICANT: WATSON, JAMES D.
APPLICANT: TAN, PAUL, L.J.
TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-
NUMBER OF SEQUENCES: 194
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman

STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/997,080
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1007
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 273 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-997-080-75

Query Match 83.9%; Score 26; DB 2; Length 273;
Best Local Similarity 66.7%; Pred. No. 3.8e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXMR 6
DB 267 FDPWMR 272

RESULT 10
US-08-997-362-75
Sequence 75, Application US/08997362
Patent No. 5985287
GENERAL INFORMATION:
APPLICANT: Tan, Paul
APPLICANT: HiYama, Jun
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Scott, Linda
APPLICANT: Prestidge, Ross
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
NUMBER OF SEQUENCES: 194
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/997,362
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970
FILING DATE: June 12, 1997
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347
FILING DATE: August 29, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002c2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 273 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-997-362-75

Query Match 83.9%; Score 26; DB 2; Length 273;
Best Local Similarity 66.7%; Pred. No. 3.8e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXMR 6
DB 267 FDPWMR 272

RESULT 11
US-08-873-970-75
Sequence 75, Application US/08873970
Patent No. 6001361
GENERAL INFORMATION:
APPLICANT: Tan, Paul
APPLICANT: HiYama, Jun
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Scott, Linda
APPLICANT: Prestidge, Ross
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/873,970
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/705,347
FILING DATE: 29-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563

TELEX:
: INFORMATION FOR SEQ ID NO: 75:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 273 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-873-970-75

Query Match : 83.9%; Score 26; DB 3; Length 273;
Best Local Similarity 66.7%; Pred. No. 3.8e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPXMR 6
DB 267 FDPWMR 272

RESULT 12

US-09-095-855-75
: Sequence 75, Application US/09095855
: Patent No. 6160093

GENERAL INFORMATION:

APPLICANT: Tan, Paul

APPLICANT: Visser, Elizabeth

APPLICANT: Skinner, Margot

APPLICANT: Prestidge, Ross

TITLE OF INVENTION: Compounds and Methods for
: TITLE OF INVENTION: Treatment and Diagnosis of Mycobacterial Infections

NUMBER OF SEQUENCES: 208

CORRESPONDENCE ADDRESS:

ADDRESSEE: Law Offices of Ann W. Speckman

STREET: 2601 Elliott Avenue, Suite 4185

CITY: Seattle

STATE: WA

COUNTRY: USA

ZIP: 98121

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/095,855

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/705,347

FILING DATE: 29-AUG-1996

APPLICATION NUMBER: 08/873,970

FILING DATE: 12-JUN-1997

APPLICATION NUMBER: 08/997,362

FILING DATE: 23-DEC-1997

ATTORNEY/AGENT INFORMATION:

NAME: Sleath, Janet

REGISTRATION NUMBER: 37,007

REFERENCE/DOCKET NUMBER: 11000.1002c3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-269-0565

TELEFAX: 206-269-0563

TELEX:

INFORMATION FOR SEQ ID NO: 75:

SEQUENCE CHARACTERISTICS:

LENGTH: 273 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

Query Match : 83.9%; Score 26; DB 4; Length 273;
Best Local Similarity 66.7%; Pred. No. 3.8e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPXMR 6
DB 267 FDPWMR 272

RESULT 13

US-08-997-080-194

: Sequence 194, Application US/08997080

: Patent No. 5968524

GENERAL INFORMATION:

APPLICANT: WATSON, JAMES D.

APPLICANT: TAN, PAUL L.J.

TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-

NUMBER OF SEQUENCES: 194

CORRESPONDENCE ADDRESS:

ADDRESSEE: Law Offices of Ann W. Speckman

STREET: 2601 Elliott Avenue, Suite 4185

CITY: Seattle

STATE: WA

COUNTRY: USA

ZIP: 98121

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/997,080

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Sleath, Janet

REGISTRATION NUMBER: 37,007

REFERENCE/DOCKET NUMBER: 11000.1007

TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-269-0565

TELEFAX: 206-269-0563

TELEX:

INFORMATION FOR SEQ ID NO: 194:

SEQUENCE CHARACTERISTICS:

LENGTH: 370 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-997-080-194

Query Match : 83.9%; Score 26; DB 2; Length 370;
Best Local Similarity 66.7%; Pred. No. 5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 EXPXMR 6
DB 364 FDPWMR 369

RESULT 14

US-08-997-362-194

: Sequence 194, Application US/08997362

: Patent No. 5985287

GENERAL INFORMATION:

APPLICANT: Tan, Paul

APPLICANT: Miyama, Jun

APPLICANT: Visser, Elizabeth

APPLICANT: Skinner, Margot

APPLICANT: Scott, Linda

APPLICANT: Prestidge, Ross
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
NUMBER OF SEQUENCES: 194
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER-READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/997,362
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970
FILING DATE: June 12, 1997
APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347
FILING DATE: August 29, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002c2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ ID NO: 194:
SEQUENCE CHARACTERISTICS:
LENGTH: 370 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-997-362-194

Query Match 83.9% Score 26; DB 2; Length 370;
Best Local Similarity 66.7% Pred. NO. 5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FXPXWR 6
1 1 1 1
Db 364 FDPWWR 369

RESULT 15
US-09-095-855-194
Sequence 194, Application US/09095855
Patent No. 6160093
GENERAL INFORMATION:
APPLICANT: Tan, Paul
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Prestidge, Ross
TITLE OF INVENTION: Compounds and Methods for
TITLE OF INVENTION: Treatment and Diagnosis of Mycobacterial Infections
NUMBER OF SEQUENCES: 208
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER-READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/095,855
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/705,347
FILING DATE: 29-AUG-1996
APPLICATION NUMBER: 08/873,970
FILING DATE: 12-JUN-1997
APPLICATION NUMBER: 08/997,362
FILING DATE: 23-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002c3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ ID NO: 194:
SEQUENCE CHARACTERISTICS:
LENGTH: 370 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-095-855-194

QY 1 FXPXWR 6
1 1 1 1
Db 364 FDPWWR 369

Search completed: February 27, 2002, 11:36:08
Job time: 147 sec

Wed Feb 27 12:13:57 2002

us-09-446-109a-18.ra1

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:41:20 : Search time 303.5 Seconds
(without alignments)
1.464 Million cell updates/sec

Title: US-09-446-109a-19
Sequence: 1 FWPXMR 6

Scoring table: BLUSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAT:
2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAT:
3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAT:
4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAT:
5: /SID58/gcgdata/geneseq/geneseq/AA1984.DAT:
6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT:
7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAT:
8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAT:
9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAT:
10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAT:
11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAT:
12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAT:
13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAT:
14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAT:
15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAT:
16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT:
17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAT:
18: /SID58/gcgdata/geneseq/geneseq/AA1997.DAT:
19: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT:
20: /SID58/gcgdata/geneseq/geneseq/AA1999.DAT:
21: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT:
22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT:

Prod. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	95.1	262	21	Arabidopsis thaliana
2	39	95.1	282	21	Arabidopsis thaliana
3	39	95.1	283	21	Arabidopsis thaliana
4	35	85.4	39	21	Arabidopsis thaliana
5	35	85.4	86	21	Arabidopsis thaliana
6	35	85.4	3165	14	Human secreted protein
7	34	82.9	10	22	Human protein frag
8	34	82.9	46	22	Peptide #3108 enco
9	34	82.9	46	22	Peptide #3105 enco
10	34	82.9	46	22	Peptide #2979 enco
11	34	82.9	54	21	Human secreted pro

12	34	82.9	55	21	AA603214	Human secreted pro
13	34	82.9	59	20	AA725774	Human secreted pro
14	34	82.9	61	22	AA675547	Human colon cancer
15	34	82.9	76	21	AA843272	Human ORFX ORF3036
16	34	82.9	78	21	AA602532	Human secreted pro
17	34	82.9	80	22	AA674485	Human colon cancer
18	34	82.9	88	21	AA839484	Human secreted pro
19	34	82.9	90	21	AA840630	Human ORFX ORF394
20	34	82.9	95	22	AA813768	Peptide #202 enco
21	34	82.9	95	22	AA826166	Peptide #203 enco
22	34	82.9	95	22	AA801513	Peptide #195 enco
23	34	82.9	101	20	AA729190	Amino acid sequenc
24	34	82.9	123	21	AA759922	Rat skin cell prot
25	34	82.9	123	21	AA855931	Skin cell protein,
26	34	82.9	138	21	AA842364	Human ORFX ORF2148
27	34	82.9	159	22	AA833640	Human protein sequ
28	34	82.9	290	21	AA78511	Human uncoupling p
29	34	82.9	291	21	AA790322	Novel human protei
30	34	82.9	322	21	AA794666	Human uncoupling p
31	34	82.9	322	21	AA794668	Murine uncoupling
32	34	82.9	325	21	AA794665	Human uncoupling p
33	34	82.9	325	21	AA794669	Murine uncoupling
34	34	82.9	334	22	AA606056	Human gene 16 enco
35	34	82.9	335	21	AA787079	Human secreted pro
36	34	82.9	353	21	AA794667	Human uncoupling p
37	34	82.9	356	22	AA840377	Human polypeptide
38	34	82.9	356	22	AA847146	CD1F-28, incycle I
39	34	82.9	539	19	AA662301	Splnach D1 proteas
40	34	82.9	620	21	AA632089	Arabidopsis thaliana
41	34	82.9	649	21	AA632088	Arabidopsis thaliana
42	34	82.9	656	21	AA651804	Arabidopsis thaliana
43	34	82.9	664	21	AA632092	Arabidopsis thaliana
44	34	82.9	685	21	AA651803	Arabidopsis thaliana
45	34	82.9	686	21	AA632087	Arabidopsis thaliana

ALIGNMENTS

RESULT 1	
AA645062	standard: Protein; 262 AA.
ID	AA645062
AC	AA645062
XX	
XX	18-OCT-2000 (first entry)
DT	
XX	Arabidopsis thaliana protein fragment SEQ ID NO: 56524.
DE	
XX	Protein identification: signal transduction pathway; metabolic pathway;
KW	hybridisation assay; genetic mapping; gene expression control; promoter;
KW	termination sequence.
XX	
GS	Arabidopsis thaliana.
XX	
FN	EP1033405-A2.
XX	
PD	06-SEP-2000.
XX	
PF	25-FEB-2000; 2000EP-0301439.
XX	
FR	25-FEB-1999; 9905-0121825.
PR	05-MAR-1999; 9905-0123180.
PR	09-MAR-1999; 9905-0123548.
PR	23-MAR-1999; 9905-0125788.
PR	25-MAR-1999; 9905-0126264.
PR	29-MAR-1999; 9905-0126785.
PR	01-APR-1999; 9905-0127462.
PR	06-APR-1999; 9905-0128234.
PR	08-APR-1999; 9905-0128714.
PR	16-APR-1999; 9905-0129845.
PR	19-APR-1999; 9905-0130077.
PR	21-APR-1999; 9905-0130449.

PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131049.
PR 30-APR-1999; 99US-0131248.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139494.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142927.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145912.
PR 27-JUL-1999; 99US-0145919.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147302.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0151066.
PR 23-AUG-1999; 99US-0151080.
PR 23-AUG-1999; 99US-0151307.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151307.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156566.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.

PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 95.1%; Score 39; DB 21; Length 262;
Best Local Similarity 83.3%; Pred. No. 17;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FWPXWR 6
Db 98 fwpkwr 103

RESULT 2

ID AAG45061 standard: Protein: 282 AA.

AC AAG45061;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 56523.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

XX EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.

FR 06-MAY-1999; 99US-0132486.
FR 06-MAY-1999; 99US-0132487.
FR 07-MAY-1999; 99US-0132863.
FR 11-MAY-1999; 99US-0134256.
FR 14-MAY-1999; 99US-0134218.
FR 14-MAY-1999; 99US-0134219.
FR 14-MAY-1999; 99US-0134221.
FR 14-MAY-1999; 99US-0134370.
FR 18-MAY-1999; 99US-0134768.
FR 19-MAY-1999; 99US-0134941.
FR 20-MAY-1999; 99US-0135124.
FR 21-MAY-1999; 99US-0135353.
FR 24-MAY-1999; 99US-0135629.
FR 25-MAY-1999; 99US-0136021.
FR 27-MAY-1999; 99US-0136392.
FR 28-MAY-1999; 99US-0136782.
FR 01-JUN-1999; 99US-0137222.
FR 03-JUN-1999; 99US-0137528.
FR 04-JUN-1999; 99US-0137502.
FR 07-JUN-1999; 99US-0137724.
FR 08-JUN-1999; 99US-0138094.
FR 10-JUN-1999; 99US-0138540.
FR 10-JUN-1999; 99US-0138847.
FR 14-JUN-1999; 99US-0139119.
FR 16-JUN-1999; 99US-0139452.
FR 16-JUN-1999; 99US-0139453.
FR 17-JUN-1999; 99US-0139454.
FR 18-JUN-1999; 99US-0139454.
FR 18-JUN-1999; 99US-0139455.
FR 18-JUN-1999; 99US-0139456.
FR 18-JUN-1999; 99US-0139457.
FR 18-JUN-1999; 99US-0139458.
FR 18-JUN-1999; 99US-0139459.
FR 18-JUN-1999; 99US-0139460.
FR 18-JUN-1999; 99US-0139461.
FR 18-JUN-1999; 99US-0139462.
FR 18-JUN-1999; 99US-0139463.
FR 18-JUN-1999; 99US-0139750.
FR 21-JUN-1999; 99US-0139763.
FR 21-JUN-1999; 99US-0139817.
FR 22-JUN-1999; 99US-0139899.
FR 23-JUN-1999; 99US-0140353.
FR 23-JUN-1999; 99US-0140354.
FR 24-JUN-1999; 99US-0140695.
FR 28-JUN-1999; 99US-0140823.
FR 29-JUN-1999; 99US-0140991.
FR 30-JUN-1999; 99US-0141287.
FR 01-JUL-1999; 99US-0141842.
FR 01-JUL-1999; 99US-0142154.
FR 02-JUL-1999; 99US-0142055.
FR 06-JUL-1999; 99US-0142390.
FR 08-JUL-1999; 99US-0142803.
FR 09-JUL-1999; 99US-0142920.
FR 12-JUL-1999; 99US-0142977.
FR 13-JUL-1999; 99US-0143542.
FR 14-JUL-1999; 99US-0143624.
FR 15-JUL-1999; 99US-0144005.
FR 16-JUL-1999; 99US-0144085.
FR 16-JUL-1999; 99US-0144086.
FR 19-JUL-1999; 99US-0144325.
FR 19-JUL-1999; 99US-0144331.
FR 19-JUL-1999; 99US-0144332.
FR 19-JUL-1999; 99US-0144333.
FR 19-JUL-1999; 99US-0144334.
FR 19-JUL-1999; 99US-0144334.
FR 19-JUL-1999; 99US-0144335.
FR 20-JUL-1999; 99US-0144352.
FR 20-JUL-1999; 99US-0144632.
FR 20-JUL-1999; 99US-0144884.
FR 21-JUL-1999; 99US-0144814.
FR 21-JUL-1999; 99US-0145086.
FR 21-JUL-1999; 99US-0145088.
FR 22-JUL-1999; 99US-0145085.
FR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.

PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161104.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 95.1%; Score 39; DB 21; Length 282;
Best Local Similarity 83.3%; Pred. NO. 18;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Caps 0;

Cy 1 FWPXMR 6
Db 118 fwpkwr 123

RESULT 3
AAC45060
ID AAC45060 standard; Protein: 283 AA.
XX AAC45060;

DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 56522.
XX
KW Protein identification: signal transduction pathway; metabolic pathway;
KW hydridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
OS EPI033405-A2.
PM
XX
PD
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160860.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.

PR 25-OCT-1999; 99US-01611406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 95.1%; Score 39; DB 21; Length 283;
Best Local Similarity 83.3%; Pred. No. 18;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FWPXMR 6
Db 119 fwpkwr 124

RESULT 4
AAC55430
ID AAC55430 standard; Protein: 39 AA.
XX
AC AAC55430;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 71072.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hydrolisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137228.
PR 03-JUN-1999; 99US-0137522.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140655.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.

```

PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148664.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.

```

```

PR 29-OCT-1999; 99US-0162142.
Query Match 85.4%; Score 35; DB 21; Length 39;
Best Local Similarity 80.0%; Pred. No. 12;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 FWPXW 5
    1 1 1 1
Db 33 fwpw 37

RESULT 5
AAB23607
ID AAB23607 standard; Protein; 86 AA.
XX
AC AAB23607;
XX
DT 12-JAN-2001 (first entry)
XX
DE Human secreted protein SEQ ID NO: 14.
XX
KW Human; secreted protein; cytokine; cell proliferation;
KW nutritional supplement; immune modulation; autoimmune disorder;
KW hematopoiesis regulation; tissue growth; haemostasis; inflammation.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 28..40
FT /label= signal_peptide
FT Protein 41..86
FT /label= mature_protein
XX
UN MO200049134-A1.
XX
PD 24-AUG-2000.
XX
PF 18-FEB-2000; 2000MO-US04340.
XX
PR 19-FEB-1999; 99US-0120680.
PR 23-APR-1999; 99US-0298733.
PR 17-AUG-1999; 99US-0149639.
PR 23-SEP-1999; 99US-0155686.
PR 01-OCT-1999; 99US-0157247.
PR 29-NOV-1999; 99US-0167822.
PR 29-NOV-1999; 99US-0167823.
PR 15-FEB-2000; 2000US-0298733.
XX
FA (ALPH-) ALPHAGEN INC.
XX
FT Valenzuela D, Yuan O, Hoffman H, Hall J, Rapiejko P;
XX
DR WPI: 2000-549267/50.
XX
DR N-PSDB; AAA93107.
XX
PT New secreted proteins and polynucleotides encoding them, which are
PT derived from homo sapiens, useful for therapy, diagnosis, and research,
PT as well as nutritional sources or supplements -
XX
PS Claim 23; Page 246-247; 309pp; English.
XX
CC The present sequence is the sequence of a human secreted protein. Its
CC cDNA was isolated from a foetal brain cDNA library. The proteins
CC and coding sequences of the invention can be used in the isolation of
CC similar genes and proteins. In the elucidation of their function in vivo,
CC and to treat a number of conditions. It is possible that they may have
CC uses as nutritional supplements, as cytokine or cell proliferation
CC factors, in immune modulation, where they may be used to treat immune and
CC autoimmune diseases, as haematopoiesis regulators (treating myeloid or
CC lymphoid cell deficiencies), in the promotion of tissue growth, they may
CC have chemokine or chemotactic activity, haemostatic or thrombolytic
CC activity, or anti-inflammatory activity.
XX

```

SQ Sequence 86 AA:

Query Match 85.4%; Score 35; DB 21; Length 86;
 Best Local Similarity 66.7%; Pred. No. 26;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FWPXMR 6
 : : : : :
 Db 50 WPPAWR 55

RESULT 6

ID AAR38889 standard; Protein: 3165 AA.

AC AAR38889;

DT 25-FEB-1994 (first entry)

Sequence encoded by ORF B of the long double-stranded cytoplasmic
 RNA (L-dsRNA) present in the hypovirulent strain EP713.

hypovirulent; spore; pathogenic fungus; chestnut blight; papain;
 hypovirulence associated virus; potyvirus; genetic element.

OS Cryphonectria (Endothia) parasitica strain EP713.

XX MO9316170-A.

XX 19-AUG-1993.

XX 03-FEB-1993; 93MO-US01024.

XX 06-FEB-1992; 92US-0832117.

(CHO1/) CHO1 G H.
 (NUSS/) NUSS D L.

PI CHO1 GH, NUSS DL;

DR WPI: 1993-272875/34.

DR N-PSDB; AAQ47380.

Fungal and fungal spores or components - modified to confer
 transmissible hypo-virulent phenotype, useful for controlling
 fungal diseases, e.g. chestnut blight

PS Claim 23; Fig 1; 69pp; English.

Natural strains of *C. parasitica* which are hypovirulent contain a
 cytoplasmic determinant that is transferred by hyphal anastomosis.
 The determinant is a ds RNA species which is believed to be of viral
 origin. The large dsRNA (L-dsRNA) present in hypovirulent *C.*
parasitica strain EP713 encodes two large polypeptides (AAR38888,
 AAR38889) that undergo autoproteolytic processing during translation.
 The AA sequences of these polypeptides contain five domains with
 significant similarity to conserved domains within the protein
 products encoded by members of the potyvirus gp. of +ve strand RNA
 plant viruses, and a common ancestry is implied. The term
 hypovirulence-associated virus (HAV) is used to denote this class
 of genetic element. ORF encodes two polypeptides, p29 and p40, that
 are released from a polypeptide, p69, by autocatalysis mediated by
 p29. Cleavage occurs between Gly-248 and Gly-249 during translation
 and is dependent on residues Cys-162 and His-215. Expression of ORF
 B also involves an autoproteolytic event in which a 48 kDa
 polypeptide, designated p48, is released from the N-terminal of the
 encoded polypeptide. Cleavage of p48 occurs between Gly-418 and
 Ala-419 and is dependent upon residues Cys-341 and His-388. Both
 p29 and p48 resemble papain-like proteases. Putative RNA-dependent
 RNA polymerases and RNA helicase motifs have been located in the
 C-terminal half of ORF B.

SQ Sequence 3165 AA:

Query Match 85.4%; Score 35; DB 14; Length 3165;
 Best Local Similarity 66.7%; Pred. No. 9.3e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FWPXMR 6
 : : : : :
 Db 1002 fwpdWq 1007

RESULT 7

ID AAM00436 standard; Peptide: 10 AA.

AC AAM00436;

DT 01-OCT-2001 (first entry)

Human protein fragment SEQ ID NO: 984.

Human: single nucleotide polymorphism; SNP; paternity test;
 forensic test; aberrant protein expression.

XX Homo sapiens.

XX MO200151670-A2.

XX 19-JUL-2001.

XX 05-JAN-2001; 2001MO-US00322.

XX 07-JAN-2000; 2000US-0174962.

(CURA-) CURAGEN CORP.
 Shimkets RA, Leach MD;

DR WPI: 2001-451871/48.

DR N-PSDB; AAI89553.

Isolated human polynucleotides containing single nucleotide
 polymorphisms, useful for the treatment and diagnosis of e.g. cancer,
 infection and diabetes

PS Disclosure; Page 386; 475pp; English.

The present invention relates to human nucleic acids containing single
 nucleotide polymorphisms (SNPs). These can be used in forensic and
 paternity tests, and to aid in the treatment of diseases associated with
 aberrant protein expression, including cancer, amyloidosis, diabetes,
 Alzheimer's disease, Down's syndrome, oedema, lupus (SLE), vasculitis,
 glomerulonephritis, haemolytic anaemia, thrombocytopaenia, arthritis,
 meningitis, muscular disorders, dementia, neurological diseases, tumours
 sclerostis, male infertility, hypercalcaemia, blood pressure disorders,
 osteoporosis, pathogenic infections, hypercholesterolaemia, obesity or
 autoimmunity. The present sequence is a peptide encoded by a
 polymorphism-containing oligonucleotide fragment of the invention.

SQ Sequence 10 AA:

Query Match 82.9%; Score 34; DB 22; Length 10;
 Best Local Similarity 80.0%; Pred. No. 4.7;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FWPXW 5
 : : : : :
 Db 5 fwpwW 9

RESULT 8

AA16584
ID AA16584 standard; Protein: 46 AA.
XX
AC AA16584;
XX
XX
DT 12-OCT-2001 (first entry)
XX
XX Peptide #3018 encoded by probe for measuring cervical gene expression.
DE
XX Probe: human; microarray: gene expression; cervical epithelial cell;
KM cervical cancer;
XX
XX Homo sapiens;
OS
XX MO200157278-A2.
PN
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00670.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000CB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI: 2001-488901/53.
DR
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
XX
XX Claim 27; SEQ ID NO 21410; 487pp; English.
PS
XX
XX The present invention relates to human single exon nucleic acid probes
CC (SENP: see A110068-AA128459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human Hela cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
CC
XX
XX Sequence 46 AA:
SQ

Query Match 82.9%: Score 34; DB 22; Length 46;
Best Local Similarity 80.0%: Pred. No. 21;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 2 WPXWR 6
II II
Db 30 wpsvr 34

RESULT 9
ID AA16584 standard; Protein: 46 AA.
XX
AC AA16584;
XX
XX
DT 17-OCT-2001 (first entry)
XX
XX Peptide #3105 encoded by probe for measuring placental gene expression.
DE
XX

KM Probe: microarray; human; placenta; antenatal diagnosis;
KM genetic disorder.
XX
XX
XX Homo sapiens.
XX
XX MO200157272-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00663.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000CB-0024263.
XX
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI: 2001-488907/53.
DR
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
XX
XX Claim 27; SEQ ID NO 29337; 654pp; English.
PS
XX
XX The present invention relates to single exon nucleic acid probes (SENP:
CC see A113137-A157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
CC
XX
XX Sequence 46 AA:
SQ

Query Match 82.9%: Score 34; DB 22; Length 46;
Best Local Similarity 80.0%: Pred. No. 21;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 2 WPXWR 6
II II
Db 30 wpsvr 34

RESULT 10
ID AA16584 standard; Protein: 46 AA.
XX
AC AA16584;
XX
XX
DT 09-OCT-2001 (first entry)
XX
XX Peptide #2979 encoded by probe for measuring breast gene expression.
DE
XX
XX Probe: human; breast disease; breast cancer; development disorder;
KM inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX Homo sapiens;
OS
XX MO200157270-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 29-JAN-2001; 2001WO-US00661.
PR
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR

PR 03-AUG-2000: 2000US-0632366.
 PR 21-SEP-2000: 2000US-0234687.
 PR 27-SEP-2000: 2000US-0236359.
 PR 04-OCT-2000: 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR:
 XX WPI: 2001-476286/51.
 DR
 XX
 PT Novel single exon nucleic acid probe used to measuring gene expression
 PT in a human breast -
 XX
 PS Claim 27: SEQ ID NO 13037; 322pp; English.
 XX
 CC The present invention relates to novel single exon nucleic acid probes
 CC (see AA100010-AA110067). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for measuring human gene expression in
 CC a human breast sample, where the probe hybridises at high stringency to a
 CC nucleic acid expressed in the human breast. The probes are useful for
 CC predicting, diagnosing, grading, staging, monitoring and prognosing
 CC diseases of the human breast, particularly those diseases with polygenic
 CC aetiology. The diseases include: breast cancer, disorders of development,
 CC inflammatory diseases of the breast, fibrocystic changes, proliferative
 CC breast disease and non-carcinoma tumours.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 46 AA:

Query Match 82.9%; Score 34; DB 22; Length 46;
 Best Local Similarity 80.0%; Pred. No. 21;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 WPXMR 6
 II II
 DB 30 WPSWC 34

RESULT 11
 AAC02493
 ID AAC02493 standard; Protein: 54 AA.
 XX
 AC AAC02493:
 XX
 DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein, SEQ ID NO: 6574.
 XX
 KW Human, 5' EST: expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.
 XX
 OS Homo sapiens.
 XX
 PN EP1033401-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 21-FEB-2000: 2000EP-0200610.
 XX
 PR 26-FEB-1999: 99US-0122487.
 XX
 PA (GEST) GENSET.
 XX
 PI Dumas Milne Edwards J, Duclet A, Giordano J;
 XX
 DR WPI: 2000-500381/45.
 DR N-PSDB: AAC02499.
 XX
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for

PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 13: SEQ ID 6574; 71pp + CD-ROM; English.
 XX

CC The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30
 CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (UTR) of the mRNA because they are often obtained
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 CC those cases where longer cDNA sequences have been obtained, the full 5'
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 CC ends and can therefore be used to obtain full length cDNAs and genomic
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.
 XX

Query Match 82.9%; Score 34; DB 21; Length 54;
 Best Local Similarity 80.0%; Pred. No. 25;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 WPXMR 6
 II II
 DB 25 WPLWR 29

RESULT 12
 AAG03214
 ID AAG03214 standard; Protein: 55 AA.
 XX
 AC AAG03214:
 XX
 DT 06-OCT-2000 (first entry)
 XX
 DE Human secreted protein, SEQ ID NO: 7295.
 XX
 KW Human, 5' EST: expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.
 XX
 OS Homo sapiens.
 XX
 PN EP1033401-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 21-FEB-2000: 2000EP-0200610.
 XX
 PR 26-FEB-1999: 99US-0122487.
 XX
 PA (GEST) GENSET.
 XX
 PI Dumas Milne Edwards J, Duclet A, Giordano J;
 XX
 DR WPI: 2000-500381/45.
 DR N-PSDB: AAC03220.
 XX
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 13: SEQ ID 7295; 71pp + CD-ROM; English.
 XX
 CC The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30
 CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (UTR) of the mRNA because they are often obtained
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for

CC Isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.

XX Sequence 55 AA:

Query Match: 82.9%; Score 34; DB 21; Length 55;

Best Local Similarity 80.0%; Pred. No. 25;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 WPXMR 6

Db 25 wptwr 29

RESULT 13

AAV25774

ID AAV25774 standard; Protein: 59 AA..

XX AAV25774:

XX 04-OCT-1999 (first entry)

XX Human secreted protein encoded from gene 64.

XX Secreted protein: human; treatment; diagnosis; therapy; cancer; tumour;
XX neurodegenerative disorder; developmental abnormality; blood disorder;
XX fetal deficiency; blood disorder; leukemia; immune system; inflammation;
XX autoimmune disease; hepatic disease; renal disease; allergy; restenosis;
XX ischaemic shock; Alzheimer's disease; cognitive disorder; schizophrenia;
XX cardiovascular disorder; wound healing; stroke; arthritis; obesity;
XX asthma; sepsis; acne; psoriasis; transplant rejection; infection; AIDS;
XX metabolic disorder.

XX Homo sapiens.

XX W09938881-A1.

XX 05-AUG-1999.

XX 27-JAN-1999; 99WO-US01621.

XX 30-JAN-1998; 98US-0073170.

XX 30-JAN-1998; 98US-0073159.

XX 30-JAN-1998; 98US-0073160.

XX 30-JAN-1998; 98US-0073161.

XX 30-JAN-1998; 98US-0073162.

XX 30-JAN-1998; 98US-0073165.

XX 30-JAN-1998; 98US-0073167.

XX (HUMA-) HUMAN GENOME SCI INC.

XX WPI: 1999-469315/39.

XX N-PSDB: AA200473.

XX New isolated human genes and the secreted polypeptides they encode

XX Claim 1b: Page 336; 393pp; English.

CC This invention describes novel human genes (see AA200410-200477) and the
CC secreted proteins (see AAV25774-25778) and fragments (see
CC AAV25779-257907) they encode. The polynucleotides and their corresponding

CC secreted polypeptides are useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. Also pathological
CC conditions can be diagnosed by determining the amount of the new

CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 67
CC polynucleotides of the invention, based on which tissues they are most
CC highly expressed in, and include developing products for the diagnosis
CC or treatment of cancer, tumours, neurodegenerative disorders,
CC developmental abnormalities and fetal deficiencies, blood disorders,
CC leukemias, diseases of the immune system, autoimmune diseases, hepatic
CC and renal disease, inflammation, allergies, ischaemic shock, Alzheimer's
CC and cognitive disorders, schizophrenia, restenosis, cardiovascular
CC disorders, wound healing, stroke, arthritis, obesity, asthma, sepsis,
CC acne, psoriasis, transplant rejection, metabolic disorders, infections
CC and AIDS. The polypeptides are also useful for identifying their binding
CC partners.

XX Sequence 59 AA:

Query Match: 82.9%; Score 34; DB 20; Length 59;

Best Local Similarity 80.0%; Pred. No. 27;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 WPXMR 6

Db 37 wptwr 41

RESULT 14

AA675547

ID AAG75547 standard; Protein: 61 AA.

XX AAG75547:

XX 03-SEP-2001 (first entry)

XX Human colon cancer antigen protein SEQ ID NO:6311.

XX Human: colon cancer; colon cancer antigen; diagnosis; detection;

XX colorectal carcinoma.

XX Homo sapiens.

XX W0200122920-A2.

XX 05-APR-2001.

XX 28-SEP-2000; 2000WO-US26524.

XX 29-SEP-1999; 99US-0157137.

XX 03-NOV-1999; 99US-0163280.

XX (HUMA-) HUMAN GENOME SCI INC.

XX WPI: 2001-23537/24.

XX N-PSDB: AAH34952.

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -

XX Claim 11: Page 7772-7773; 9803pp; English.

CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions

CC In a patient's genome that affect the activity of p by expressing
 CC inactive proteins or to supplement the patients own production of p.
 CC Additionally, N may be used to produce the colon cancer-associated ps,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
 CC and AAB77789 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
 CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 CC
 XX Sequence: .j61 AA:
 SQ

QY 2 WPXWR 6
 11 11
 Db 7 wpswr 11

RESULT 15
 AAB43272
 ID AAB43272 standard: Protein: 76 AA.

AC AAB43272:
 DT 08-FEB-2001 (first entry)

DE Human ORFX ORF3036 polypeptide sequence SEQ ID NO:6072.

KW Human: open reading frame; ORFX: detection; cytostatic; hepatotropic;
 KW vulnery; antiposrotatic; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antinflammatory;
 KW antiviral; antibacterial; antifungal; antineumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;
 KW thrombosis; contraceptive.

OS Homo sapiens.

PN MO200058473-A2.

PD 05-OCT-2000.

PE 31-MAR-2000: 2000WO-US08621.

PR 31-MAR-1999: 99US-0127607.
 PR 02-APR-1999: 99US-0127636.
 PR 05-APR-1999: 99US-0127728.
 PR 30-MAR-2000: 2000US-0540763.

PA (CURA-) CURAGEN CORP.

PI Shimkets RA, Leach M:

DR WPI: 2000-602362/57.
 DR N-PSDB: AAC77481.

PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -

PS Claim 11: Page 5256: 5507pp: English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antiposrotatic; antiparkinsonian; neurotropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antinflammatory; antibacterial; antiviral; antifungal; antineumatic;
 CC antithyroid; and antianaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

SQ Sequence 76 AA:

Query Match 82.9%: Score 34: DB 21: Length 76;
 Best Local Similarity 80.0%: Pred. NO. 34;
 Matches 4: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

QY 2 WPXWR 6
 11 11
 Db 32 wpswr 36

Search completed: February 27, 2002, 11:41:20
 Job time: 459 sec

Wed Feb 27 12:13:58 2002

us-09-446-109a-19.rag

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:45:23 : Search time 145.23 Seconds
(without alignments)
3.147 Million cell updates/sec

Title: US-09-446-109A-19

Perfect score: 41

Sequence: 1 FWPXMR 6

Scoring table: BL0SUM62
Gapop 10.0 , Capext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: PIR_68:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	97.6	318	2	G82350
2	39	95.1	283	2	T48332
3	39	95.1	407	2	F70764
4	35	85.4	297	2	F64831
5	35	85.4	297	2	H85617
6	35	85.4	387	2	A83388
7	35	85.4	476	2	T29054
8	35	85.4	605	2	T15291
9	35	85.4	3165	2	S15010
10	35	82.9	113	2	T12282
11	34	82.9	129	2	A85686
12	34	82.9	148	2	B72772
13	34	82.9	153	2	A55139
14	34	82.9	261	2	F64924
15	34	82.9	261	2	E85774
16	34	82.9	301	2	G83556
17	34	82.9	325	3	JC7553
18	34	82.9	335	3	A43900
19	34	82.9	373	2	D71142
20	34	82.9	409	2	B75010
21	34	82.9	423	2	T16462
22	34	82.9	539	2	JH0263
23	34	82.9	552	2	S46978
24	34	82.9	580	2	T30583
25	34	82.9	640	2	T28631
26	34	82.9	665	1	P28P6
27	34	82.9	686	2	T08919
28	34	82.9	822	1	T0MCP
29	34	82.9	902	2	B84652

30	34	82.9	1099	1	S31926	myosin IB heavy
31	34	82.9	2796	2	JC4743	fatty-acid synthas
32	33	80.5	113	1	RKKRS2	ribulose-bisphosph
33	33	80.5	208	2	D83241	probable nucleas
34	33	80.5	255	2	S64574	probable membranc
35	33	80.5	269	2	D82060	conserved hypochet
36	33	80.5	292	2	H82214	conserved hypochet
37	33	80.5	305	2	T23022	hypothetical prote
38	33	80.5	323	2	T19681	hypothetical prote
39	33	80.5	326	2	C72483	probable formate d
40	33	80.5	336	2	T44958	hypothetical prote
41	33	80.5	342	2	PC4211	hepatocellular car
42	33	80.5	368	1	BGHUN	biglycan precursor
43	33	80.5	441	2	G83212	hypothetical prote
44	33	80.5	468	2	T35928	probable aldehyde
45	33	80.5	480	2	E75504	hypothetical prote

ALIGNMENTS

RESULT 1

lipid A biosynthesis lauroyl acyltransferase VC0213 [imported] - Vibrio cholerae (str

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C:Accession: G82350

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.

chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoti, I.; Sellers

l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen vibrio cholerae.

A:Reference number: A82035; MUID:20406833

A:Accession: G82350

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-318 <HEI>

A:Cross-references: GB:AE004111; GB:AE003852; NID:9654614; PID:NAF93389.1; GSPDB:GN

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC0213

A:Map position: 1

Query Match 97.6%, Score 40; DB 2; Length 318;
Best Local Similarity 83.3%, Pred. No. 16;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FWPXMR 6
DU 106 FWPXMR 111

RESULT 2

hypothetical protein F15A17.140 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T48332

R:Bevan, M.; Terry, N.; Airdies, W.; Buysshardt, C.; Dasseville, R.; De Clerck, R.;

ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.

submitted to the protein Sequence Database, April 2000

A:Reference number: Z24491

A:Accession: T48332

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-283 <BEV>

A:Cross-references: EMBL:AL163002

A:Experimental source: cultivar Columbia; BAC clone F15A17

C:Genetics:

A:Map position: 5

A:Note: F15A17.140

Query Match 95.1%; Score 39; DB 2; Length 283;
 Best Local Similarity 83.3%; Pred. No. 20;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FWPXMR 6
 111 11
 DB 119 FWPXMR 124

RESULT 3

hypothetical protein RV2067c - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: F70764
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellton, S.; Squares, S.; Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulistio, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98293987
 A:Accession: F70764
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-407 <COL>
 A:Cross-references: GB:Z73966; GB:AL123456; NID:q3261577; PIDN:CAA98215.1; PID:e247086;
 A:Experimental source: strain H37RV
 A:Genetics:
 A:Gene: RV2067c

Query Match 95.1%; Score 39; DB 2; Length 407;
 Best Local Similarity 83.3%; Pred. No. 29;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FWPXMR 6
 111 11
 DB 336 FWPXMR 341

RESULT 4

hypothetical protein b0919 - Escherichia coli
 C:Species: Escherichia coli
 C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
 C:Accession: F64831
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617
 A:Accession: F64831
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-297 <BLAT>
 A:Cross-references: GB:AE000194; GB:U00096; NID:q1787148; PIDN:AAC74005.1; PID:q1787149;
 A:Experimental source: strain K-12, substrain MG1655
 C:Superfamily: Escherichia coli hypothetical protein b0919

Query Match 85.4%; Score 35; DB 2; Length 297;
 Best Local Similarity 80.0%; Pred. No. 87;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FWPXW 5
 111 1
 DB 143 FWPXW 147

RESULT 5
 H85617

hypothetical protein 21265 [imported] - Escherichia coli (strain O157:H7)

C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
 C:Accession: H85617
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
 Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimanche, E.; Potamosis, K.; Apicini,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: H85617
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-297 <STO>
 A:Cross-references: GB:AE005174; NID:q12514088; PIDN:AAG55404.1; GSPDB:GN00145; UMG:
 A:Experimental source: strain O157:H7, substrain EDL933
 A:Genetics:
 A:Gene: 21265
 C:Superfamily: Escherichia coli hypothetical protein b0919

Query Match 85.4%; Score 35; DB 2; Length 297;
 Best Local Similarity 80.0%; Pred. No. 87;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FWPXW 5
 111 1
 DB 143 FWPXW 147

RESULT 6

probable MFS transporter PA2068 [imported] - Pseudomonas aeruginosa (strain PAO1)
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 23-Mar-2001
 C:Accession: A83388
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
 Adam, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidis, K.; L.
 .; Lory, S.; Olson, M.V.; Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pa
 A:Reference number: A82950; MUID:20437337
 A:Accession: A83388
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-387 <STO>
 A:Cross-references: GB:AE004633; GB:AE004091; NID:99948070; PIDN:AAG05456.1; GSPDB:GN
 A:Experimental source: strain PAO1
 C:Genetics:
 A:Gene: PA2068
 C:Superfamily: maltose permease

Query Match 85.4%; Score 35; DB 2; Length 387;
 Best Local Similarity 80.0%; Pred. No. 11e02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FWPXW 5
 111 1
 DB 18 FWPXW 22

RESULT 7

probable transmembrane transport protein - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
 C:Accession: T29054
 R:Redenbach, M.; Kleser, H.M.; Denapite, D.; Eichner, A.; Cullum, J.; Kinashi, H.; H
 Hol, Microbiol. 21, 77-96, 1996
 A:Title: A set of ordered cosmid and a detailed genetic and physical map for the 8 M
 A:Reference number: 220556; MUID:97000351
 A:Accession: T29054
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
 A:Residues: 1-476 <RED>
 A:Cross-references: EMBL:AL031124; NID:e1312893; P1D:e1312898; P1DN:CAA19972.1
 C:Genetics:
 A:Note: SCIC2.05c

Query Match 85.4%; Score 35; DB 2; Length 476;
 Best Local Similarity 80.0%; Pred. No. 1.4e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FWPXM 5
 ||| |
 Db 202 FWPXM 206

RESULT 8
 T15291
 hypothetical protein B0252.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 23-Mar-2001
 C:Accession: T15291

R:Du. 2
 Submitted to the EMBL Data Library, July 1995
 A:Description: The sequence of C. elegans cosmid B0252.
 A:Reference number: S59415

A:Accession: T15291
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA

A:Residues: 1-605 <DU2>
 A:Cross-references: EMBL:U23453; NID:g733572; P1D:g733575; P1DN:AAC46756.1; CESP:B0252.2
 A:Experimental source: strain Bristol N2
 C:Genetics:

A:Gene: CESP:B0252.2
 A:Initons: 65/1; 93/2; 132/3; 318/1; 360/2; 387/2; 444/2; 478/3; 507/3; 569/2
 C:Superfamily: acid sphingomyelinase; phosphoserine core homology; saposin repeat homology

Query Match 85.4%; Score 35; DB 2; Length 605;
 Best Local Similarity 80.0%; Pred. No. 1.7e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FWPXM 5
 ||| |
 Db 333 FWPXM 337

RESULT 9
 S15010

hypothetical protein B - Cryphonectria hypovirus 1
 C:Species: Cryphonectria hypovirus 1
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 21-Jul-2000
 C:Accession: S15010

R:Shapira, R.; Choi, G.H.; Nuss, D.L.
 FMBD J. 10, 731-739, 1991

A:Title: Virus-like genetic organization and expression strategy for a double-stranded RNA
 A:Reference number: S15009; MUID:91184117
 A:Accession: S15010

A:Status: preliminary
 A:Molecule type: genomic RNA
 A:Residues: 1-3165 <EMB>
 A:Cross-references: GB:M57938; NID:g331157; P1DN:AAA67458.1; P1D:g331159

Query Match 85.4%; Score 35; DB 2; Length 3165;
 Best Local Similarity 66.7%; Pred. No. 7.8e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FWPXM 6
 ||| |
 Db 1002 FWPXM 1007

RESULT 10
 T42282
 hypothetical 12.5k protein - phage SP1

C:Species: phage SP1
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 11-May-2000
 C:Accession: T42282

R:Alonso, J.C.; Luder, G.; Stiege, A.C.; Chai, S.; Weise, F.; Trautner, W.A.
 Gene 204, 201-212, 1997

A:Title: The complete nucleotide sequence and functional organization of Bacillus sub
 A:Reference number: 222137; MUID:98094274

A:Accession: T42282
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA

A:Residues: 1-113 <ALO>
 A:Cross-references: EMBL:X97918; P1DN:CAA65592.1
 C:Superfamily: phage SP1 hypothetical 12.5k protein

Query Match 82.9%; Score 34; DB 2; Length 113;
 Best Local Similarity 80.0%; Pred. No. 51;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FWPXM 5
 ||| |
 Db 23 FWPXM 27

RESULT 11
 A85686

probable holin protein of prophage CP-933C 21852 [imported] - Escherichia coli (strai
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
 C:Accession: A85686

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
 Hiler, L.; Grothbeck, E.J.; Davis, M.W.; Lim, A.; Diallanza, E.; Potamou, K.; Apoda
 Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: A85686
 A:Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-129 <STO>
 A:Cross-references: GB:AE005174; NID:g12514774; P1DN:AAG55949.1; GSPD:GN00145; UMGF:
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:

A:Gene: 21852

Query Match 82.9%; Score 34; DB 2; Length 129;
 Best Local Similarity 80.0%; Pred. No. 58;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FWPXM 5
 ||| |
 Db 76 FWPXM 80

RESULT 12
 R72772

hypothetical protein APE0165 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jun-2000
 C:Accession: R72772

R:Kawarabayashi, Y.; Iino, Y.; Horikawa, H.; Yamazaki, S.; Hukawa, Y.; Jin-no, K.; Ta
 kawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.
 DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, A
 A:Reference number: A72450; MUID:99310339
 A:Accession: R72772

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-148 <KMW>
 A:Cross-references: DDBJ:AP000058; NID:g5103388; P1DN:BAAG79076.1; P1D:d1042852; P1D:g

A:Experimental source: strain K1
 C:Genetics:
 A:Gene: APE0165
 C:Superfamily: Aeropyrum pernix hypothetical protein APE0165

Query Match 82.9% Score 34 DB 2 Length 148;
 Best Local Similarity 80.0% Pred. No. 65;
 Matches 4: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 WPXW 6
 DB 42 WRMR 46

RESULT 13

A55139
 myoglobin, body wall - pig roundworm
 C:Species: Ascaris suum (pig roundworm)
 C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 04-Mar-2000
 C:Accession: A55139
 R:Blaxter, M.L.; Vanfleteren, J.R.; Xia, J.; Moens, L.
 J. Biol. Chem. 269, 30181-30186, 1994
 A:Title: Structural characterization of an Ascaris myoglobin.
 A:Reference number: A55139; MUID:95074010
 A:Accession: A55139
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-153 <BLA>
 A:Cross-references: GB:017337; NID:9596073; PIDN:AAA64695.1; PID:9596074
 C:Genetics:
 A:Introns: #status absent
 C:Superfamily: globin; globin homology
 C:Keywords: blocked amino end; chromoprotein; dimer; heme; iron; metalloprotein; oxygen
 F:2.147/Domain: globin homology <GB>
 F:62/Binding site: oxygen (Gln) (distal axial ligand) #status predicted
 F:94/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match 82.9% Score 34 DB 2 Length 153;
 Best Local Similarity 80.0% Pred. No. 67;
 Matches 4: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FWPXW 5
 DB 109 FWPXW 113

RESULT 14

F64924
 probable thiosulfate--dithiol sulfurtransferase (EC 2.8.1.5) - Escherichia coli
 C:Species: Escherichia coli
 C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 19-May-2000
 C:Accession: F64924
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CC
 A.: Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617
 A:Accession: F64924
 A:Molecule type: DNA
 A:Residues: 1-261 <BLAT>
 A:Cross-references: GB:AE000262; GB:U00096; NID:q1787955; PIDN:AACT4740.1; PID:q1787959;
 A:Experimental source: strain K-12, substrain MG1655
 C:Superfamily: hyc protein
 C:Keywords: sulfurtransferase; transmembrane protein
 F:30-46/Domain: transmembrane #status predicted <TM01>
 F:83-99/Domain: transmembrane #status predicted <TM02>
 F:112-128/Domain: transmembrane #status predicted <TM03>
 F:187-203/Domain: transmembrane #status predicted <TM04>
 F:224-240/Domain: transmembrane #status predicted <TM05>

Query Match 82.9% Score 34 DB 2 Length 261;
 Best Local Similarity 80.0% Pred. No. 1.1e+02;
 Matches 4: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FWPXW 5
 DB 25 FWPXW 29

RESULT 15

E85774
 hypothetical protein ydhv [imported] - Escherichia coli (strain O157:H7)
 C:Species: Escherichia coli
 C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
 C:Accession: E85774
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
 Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: E85774
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-261 <STO>
 A:Cross-references: GB:AE005174; NID:q12515669; PIDN:AAG56657.1; GSPDB:GN00145; UNKCP:
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: ydhv

Query Match 82.9% Score 34 DB 2 Length 261;
 Best Local Similarity 80.0% Pred. No. 1.1e+02;
 Matches 4: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FWPXW 5
 DB 25 FWPXW 29

Search completed: February 27, 2002, 11:45:25
 Job time: 704 sec

Wed Feb 27 12:13:59 2002

us-09-446-109a-19.rpt

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:42:49 ; Search time 78.39 Seconds
(without alignments)
2.806 Million cell updates/sec

Title: US-09-446-109A-19

Perfect score: 41

Sequence: 1 FWPXMR 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	95.1	407	YK67_MYCTU	Q10678 mycobacteri
2	35	85.4	297	YCB1_ECOLI	P75845 escherichia
3	34	82.9	153	GLB2_NSCSU	P49672 ascaris suu
4	34	82.9	261	PHSC_ECOLI	P77409 escherichia
5	34	82.9	322	BMCP_MOUSE	Q92422 mus musculu
6	34	82.9	325	BMCP_HUMAN	O95258 mus sapien
7	34	82.9	335	SM34_LYPTI	O05904 lytechinus
8	34	82.9	640	YACD_RHISN	P55386 rhizobium s
9	34	82.9	664	VP2_BPPHG	P11124 bacterioph
10	34	82.9	822	CAD3_MOUSE	P10287 mus musculu
11	33	80.5	113	RBS2_CHRYI	P22860 chromallum
12	33	80.5	255	SOLA_YEAST	P53315 saccharomyc
13	33	80.5	368	PGS1_HUMAN	P21810 homo sapien
14	33	80.5	483	MSNF_BACSU	P42437 bacillus su
15	33	80.5	782	MYLF_HUMAN	O00160 homo sapien
16	33	80.5	847	MDOH_ECOLI	P73137 escherichia
17	33	80.5	1099	MYLF_MOUSE	P70248 mus musculu
18	33	80.5	2569	LMA3_MOUSE	O61789 mus musculu
19	32	78.0	103	RT14_CAEBL	P49391 caenorhabd
20	31	75.6	254	PHSC_SALTY	P37602 salmoneilla
21	31	75.6	274	YOHM_ECOLI	P76425 escherichia
22	31	75.6	1776	SECN_HUMAN	P76038 homo sapien
23	31	75.6	550	YLOO_MYCTU	O10709 mycobacteri
24	31	75.6	762	MNN1_YEAST	P39106 saccharomyc
25	31	75.6	1121	MDR6_HUMAN	O9m45 homo sapien
26	31	75.6	1221	V143_NPVAC	P24307 autographa
27	31	75.6	1223	V143_NPVAC	O83950 oryglia pseu
28	31	75.6	1776	POLR_OYV	P20127 ononis yell
29	30	73.2	165	LYCV_BPP2	P51771 bacterioph
30	30	73.2	287	YCY2_ASTIU	P34776 asclasia lon
31	30	73.2	399	FHE_YEAST	P39676 saccharomyc
32	30	73.2	417	YOR2_MITGO	O33369 neisseria g
33	30	73.2	459	RBL2_RHOSH	P29278 rhodobacter

34	30	73.2	530	1	TACY_LISSE	P31830 listeria se
35	30	73.2	557	1	FLO4_CAEBL	Q21434 caenorhabd
36	30	73.2	598	1	FRDA_PROVU	P20922 proteus vul
37	30	73.2	601	1	FRDA_ECOLI	P00363 escherichia
38	30	73.2	605	1	RTK2_GEOCY	P42159 geodia cydo
39	30	73.2	611	1	PHSC_RHIME	P50176 r poly-beta
40	30	73.2	636	1	PHSC_RHIRT	O52728 r poly-beta
41	30	73.2	718	1	YCCS_HAFIN	P44289 haemophilus
42	30	73.2	803	1	FSPO_XENLA	P35447 xenopus lae
43	30	73.2	804	1	YBPP_ECOLI	P77504 escherichia
44	30	73.2	807	1	FSPO_RAT	P35446 rattus norv
45	30	73.2	813	1	PHSC_CHLMU	Q9pk6 cilamydia m

ALIGNMENTS

```

RESULT 1
YK67_MYCTU          STANDARD:      PRT:   407 AA.
ID YK67_MYCTU
AC Q10678:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DN HYPOTHETICAL 45.9 KDA PROTEIN RV2067C.
GN RV2067C OR MT2127 OR MTCY49.06C.
OS Mycobacterium tuberculosis.
OC Bacteria: Firmicutes: Actinobacteria: Actinobacteridae:
OC Actinomycetales: Corynebacteriaceae: Mycobacteriaceae: Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV:
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekaia F.,
RA Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares R.,
RA Sultson J.E., Taylor K., Whitehead S., Barrett B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
PN [2]
PP SEQUENCE FROM N.A.
PC STRAIN-CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Biswal W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RT Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
XL -1- SIMILARITY: SOME. TO SYNECHOCOCUS STRAIN PCC 6301 ATP SYNTHASE
SUBUNITS REGION ORF 1 (AC P08442).
CC *****
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC *****
CC EMBL: 273966; CAA98215.1;
CC EMBL: AE007063; AAK46407.1;
CC TIGR: MT2127;
CC TubercuList: RV2067C;
CC Hypothetical protein: Complete proteome.
CC CONFLICT 288 E -> K (IN REF. 2).

```

SQ SEQUENCE 407 AA: 45930 MW: D9764E968439F435 CRC64:

Query Match 95.1% Score 39: DB 1: Length 407:

Best Local Similarity 83.3% Pred. No. 13:

Matches 5: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

QY 1 FWPXMR 6

DB 336 FWPXMR 341

RESULT 2

YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

AC YCB1_ECOLI

DR HSSP: P28116; IASH.
 DR InterPro: IPR000971; Globin.
 DR Pfam: PF00042; Globin; 1.
 DR PROSITE: PS01033; Globin; 1.
 KW Heme: Oxygen transport; Respiratory protein.
 FT METAL 94 94 IRON (HEME PROXIMAL LIGAND)
 FT (BY SIMILARITY).
 SO SEQUENCE 153 AA; 17454 MW; 1B3EP9A15B49B98 CRC64;

Query Match 82.9%; Score 34; DB 1; Length 153;
 Best Local Similarity 80.0%; Pred. No. 32;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FMPXW 5
 1111
 DB 109 FMPXW 113

RESULT 4
 PHSC_ECOLI
 ID PHSC_ECOLI STANDARD: PRT: 261 AA.
 AC P77403;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PHSC PROTEIN HOMOLOG.
 CN YDHU OR B1670.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE-97426617; PubMed-9278503;
 RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12."
 RL Science 277:1453-1474(1997).
 RN [12]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE-97175536; PubMed-9023191;
 RA Hensel M., Shea J.E., Beaumler A.J., Gleason C., Blatter F.R.,
 RA Holden D.W.;
 RT "Analysis of the boundaries of Salmonella pathogenicity island 2 and
 RT the corresponding chromosomal region of Escherichia coli K-12."
 RL J. Bacteriol. 179:1105-1111(1997).
 RN [31]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE-97251357; PubMed-9097039;
 RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
 RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,
 RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
 RA Ohtomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
 RA Oshima T., Saito N., Sampei G., Seki Y., Sivasubraman S.,
 RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
 RA Yamamoto Y., Horinouchi T.;
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 28.0-40.1 min region on the linkage map."
 RL DNA Res. 3:363-377(1996).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
 CC (POTENTIAL).
 CC -1- SIMILARITY: TO S.TYPHIMURIUM PHSC.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE000262; AAC74740.1;
 DR EMBL: U68703; AAB47946.1;
 DR EMBL: D90810; BAA15442.1;
 DR Ecogen: EGI3955; ydhU.
 KW Transmembrane; Inner membrane; Complete proteome.
 FT TRANSMEM 25 45 POTENTIAL.
 FT TRANSMEM 81 101 POTENTIAL.
 FT TRANSMEM 108 128 POTENTIAL.
 FT TRANSMEM 182 202 POTENTIAL.
 FT TRANSMEM 224 244 POTENTIAL.
 SO SEQUENCE 261 AA; 29583 MW; 65CF1A45691A0AF3 CRC64;

Query Match 82.9%; Score 34; DB 1; Length 261;
 Best Local Similarity 80.0%; Pred. No. 51;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FMPXW 5
 1111
 DB 25 FMPXW 29

RESULT 5
 LMCP_MOUSE
 ID BMCP_MOUSE STANDARD: PRT: 322 AA.
 AC O922B2;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE BRAIN MITOCHONDRIAL CARRIER PROTEIN-1 (BMCP-1) (SOLUTE CARRIER FAMILY
 DE 25, MEMBER 14).
 CN SLC25A14 OR BMCP1;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eultheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/C; TISSUE=Brain;
 RX MEDLINE-99069464; PubMed-9852133;
 RA Sanchez D., Fleury C., Chomikl N., Goubert M., Huang O., Neveleva M.,
 RA Gregoire F., Fastick J., Raimbault S., Levi-Meyrueis C., Mitoux B.,
 RA Collins S., Scidm M., Richard D., Warden C., Bouillard F.,
 RA Ricquier D.;
 RT "BMCP1, a novel mitochondrial carrier with high expression in the
 RT central nervous system of humans and rodents, and respiration
 RT uncoupling activity in recombinant yeast."
 RL J. Biol. Chem. 273:34611-34615(1998).
 CC -1- FUNCTION: PARTICIPATES TO THE MITOCHONDRIAL PROTON LEAK MEASURED
 CC IN BRAIN MITOCHONDRIA.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL.
 CC TISSUE SPECIFICITY: MAINLY EXPRESSED IN BRAIN, PARTICULARLY
 CC ABUNDANT IN CORTEX, HIPPOCAMPUS THALAMUS, AMYGALA AND
 CC HYPOTHALAMUS. EXPRESSED IN OTHER TISSUES TO A LESSER EXTENT.
 CC -1- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE MITOCHONDRIAL CARRIER FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF076981; AAD03674.1;
 DR MGD: MGI:1330823; SLC25A14.
 DR InterPro: IPR002067; MLC_carrier.

DR InterPro: IPR002030: Mit_uncoupling.
 DR InterPro: IPR001993: Mitoch_carrier.
 DR Pfam: PF00153: mito_carr. 3.
 DR PRINTS: PR00784: MTUNCouPLING.
 DR PRINTS: PR00926: MITOCARRIER.
 DR PRINTS: PR00927: ADPTNSLASE.
 DR PROSITE: PS00215: MITOCH_CARRIER. 3.
 DR Mitochondrion: Repeat: Transmembrane, Transport.
 KW TRANSMEM 2 22 POTENTIAL.
 FT TRANSMEM 40 60 POTENTIAL.
 FT TRANSMEM 138 158 POTENTIAL.
 FT TRANSMEM 222 242 POTENTIAL.
 FT TRANSMEM 292 312 POTENTIAL.
 SQ SEQUENCE 322 AA: 36042 MW: 83967BCCD6BDB68 CRC64:

Query Match Best Local Similarity 82.9%; Score 34; DB 1; Length 322;
 Matches 4: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FWPXW 5
 1111
 Db 296 FWPXW 300

RESULT 6
 BMCP_HUMAN STANDARD: PRT: 325 AA.
 AC 095258;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE BRAIN MITOCHONDRIAL CARRIER PROTEIN-1 (BMCP-1) (SOLUTE CARRIER FAMILY
 DE 25, MEMBER 14).
 GN SLC25A14 OR BMCP1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN 11
 RP MEDLINE=99069464; PubMed=9852133;
 RA Sanchis D., Fleury C., Chomik N., Goubert M., Huang O., Neverova M.,
 RA Gregoire F., Baslick J., Raimbault S., Levi-Meyrueis C., Mitoux B.,
 RA Collins S., Seldin M., Richard D., Warden C., Boulland F.,
 RA "Riquier D.;
 RT "BMCP1, a novel mitochondrial carrier with high expression in the
 RT central nervous system of humans and rodents, and respiration
 RT uncoupling activity in recombinant yeast.";
 RL J. Biol. Chem. 273:34611-34615(1998).
 RN 12
 RP SEQUENCE FROM N.A.
 RA Pearce A.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: PARTICIPATES TO THE MITOCHONDRIAL PROTON LEAK MEASURED
 CC IN BRAIN MITOCHONDRIA.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 CC INNER MEMBRANE (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: MAINLY EXPRESSED IN BRAIN. SOME EXPRESSION WAS
 CC NOTICED IN TESTIS AND PITUITARY.
 CC -1- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE MITOCHONDRIAL CARRIER FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: AF078544; AAD04346.1;
 CC EMBL: AL035423; CAB81251.1; -;

DR MM: 300242;
 DR InterPro: IPR002067: Mit_carrier.
 DR InterPro: IPR002030: Mit_uncoupling.
 DR InterPro: IPR001993: Mitoch_carrier.
 DR Pfam: PF00153: mito_carr. 3.
 DR PRINTS: PR00784: MTUNCouPLING.
 DR PRINTS: PR00926: MITOCARRIER.
 DR PRINTS: PR00927: ADPTNSLASE.
 DR PROSITE: PS00215: MITOCH_CARRIER. 3.
 DR Mitochondrion: Repeat: Transmembrane, Transport.
 KW TRANSMEM 3 23 POTENTIAL.
 FT TRANSMEM 43 63 POTENTIAL.
 FT TRANSMEM 141 161 POTENTIAL.
 FT TRANSMEM 225 245 POTENTIAL.
 FT TRANSMEM 295 315 POTENTIAL.
 SQ SEQUENCE 325 AA: 36201 MW: 0447EBE3B5374982 CRC64:

Query Match Best Local Similarity 82.9%; Score 34; DB 1; Length 325;
 Matches 4: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FWPXW 5
 1111
 Db 299 FWPXW 303

RESULT 7
 SM34_LYTPI STANDARD: PRT: 335 AA.
 AC 005904;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE 34 KDA SPICULE MATRIX PROTEIN PRECURSOR (LSM34).
 DE Lytechinus pictus (Painted sea urchin).
 OS Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinoidae; Echinoidea; Echinacea; Temnopleuroidea; Toxopneustidae;
 OC Lytechinus.
 OX NCBI_TaxID=7653;
 RN 11
 RP SEQUENCE FROM N.A.
 RC TISSUE-Gastrula;
 RA MEDLINE=92077276; PubMed=1743395;
 RA Livingston B.T., Shaw R., Bailey A., Wilt F.;
 RT "Characterization of a cDNA encoding a protein involved in formation
 RT of the skeleton during development of the sea urchin Lytechinus
 RT pictus.";
 RL Dev. Biol. 148:473-480(1991).
 RN 12
 RP -1- FUNCTION: MAJOR MATRIX PROTEIN OF THE SEA URCHIN EMBRYO SPICULE
 RP WHICH DIRECTS CRYSTAL GROWTH IN CERTAIN ORIENTATIONS AND INHIBIT
 RP GROWTH IN OTHERS.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- TISSUE SPECIFICITY: EMBRYO SPICULE.
 CC -1- DOMAIN: THE REPETITIVE DOMAIN MAY PROVIDE A CALCIUM BINDING
 CC MATRIX.
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: X59616; CAA42179.1;
 CC InterPro: IPR001304; lectin_c.
 CC SMART: SM00034; CLECT. 1.
 CC PROSITE: PS00615; C-TYPE_LECTIN_1; FALSE_NEG.
 CC PROSITE: PS00041; C-TYPE_LECTIN_2; 1.
 KW Matrix protein; Signal; Repeat.

FT SIGNAL 1 17 POTENTIAL.
 FT CHAIN 18 335 34 KDA SPICULE MATRIX PROTEIN.
 FT DOMAIN 7 2 C-TYPE LECTIN.
 FT DOMAIN 131 178 PRO-RICH.
 SO SEQUENCE 335 AA: 35426 MW: AA032281DDBDAE97 CRC64:

Query Match 82.9%; Score 34; DB 1; Length 335;
 Best Local Similarity 80.0%; Pred. No. 64;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 WPXMR 6
 1111
 DB 130 WPMWR 134

RESULT 8
 YACD_RHISN STANDARD: PRT: 640 AA.
 ID YACD_RHISN STANDARD: PRT: 640 AA.
 AC P55386:

DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHETICAL 71.6 KDA PROTEIN YACD.
 CN YACD.

OS Rhizobium sp. (strain NCR234).
 OC Plasmid sym PNGR234.
 OC Bacteria: Proteobacteria: alpha subdivision: Rhizobiaceae group;
 OC Rhizobiaceae: Rhizobium.
 OX NCBI_TaxID=394;
 RN 111

RP MEDLINE=97305956; PubMed=9163424;
 RA Freilberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
 RA Perret X.;
 RT "Molecular basis of symbiosis between Rhizobium and legumes.";
 RL Nature 387:394-401(1997).
 CC -1- SIMILARITY: NONE OBVIOUS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: AE000067; AAB91634.1; -;
 DR InterPro: IPR000157; TIR.
 DR SMART: SM00255; TIR: 1.
 KW Hypothetical protein; Plasmid.
 SQ SEQUENCE 640 AA: 71609 MW: 829BF90C595A3C0E CRC64:

Query Match 82.9%; Score 34; DB 1; Length 640;
 Best Local Similarity 80.0%; Pred. No. 1.1e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FWPXW 5
 1111
 DB 520 FWPFW 524

RESULT 9
 VP2_APPH6
 ID VP2_APPH6 STANDARD: PRT: 664 AA.
 AC P11124:

DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 01-JAN-1990 (Rel. 13, Last annotation update)
 DE P2 PROTEIN.
 GN P2.
 OS Bacteriophage phi-6.

CC Viruses; dsRNA viruses: Cystoviridae: Cystovirus.
 CX NCBI_TaxID=10879;
 EN 11
 SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.

RX MEDLINE=88155752; PubMed=3346944;
 RA Mindich L., Nussbaumer L., Gottlieb P., Romanschuk M., Carton J.,
 RA Frucht S., Strissman J., Bamford D.H., Kalkkinen N.;
 RT Nucleotide sequence of the large double-stranded RNA segment of
 RT bacteriophage phi 6: genes specifying the viral replicase and
 RT transcriptase.";
 RL J. Virol. 62:1180-1185(1988).

CC -1- FUNCTION: P2 IS ONE OF THE 4 STRUCTURAL PROTEINS OF THE POLYHEDRAL
 CC PROCAPSID, WHICH IS RESPONSIBLE FOR GENOMIC REPLICATION AND
 CC TRANSCRIPTION.
 CC -1- SUBUNIT: THE PROCAPSID IS COMPOSED OF 120 COPIES EACH OF P1 AND
 CC P4, ABOUT 20 COPIES OF P2, AND BETWEEN 80 AND 100 COPIES OF P7.
 CC -1- MISCELLANEOUS: P2 IS PRODUCED AT ABOUT 10% OF THE RATE OF THE
 CC OTHER PROCAPSID PROTEINS P1, P4 AND P7.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: M17461; AAA32355.1; -;
 DR PIR: B29885; P2BPF6.
 NR Transcription: Structural protein.
 KM INIT_MET 0
 PT SEQUENCE 664 AA: 74791 MW: CE2791F44CB2B58F CRC64;

Query Match 82.9%; Score 34; DB 1; Length 664;
 Best Local Similarity 80.0%; Pred. No. 1.2e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FWPXW 5
 1111
 DB 331 FWPFW 335

RESULT 10
 CAD3_MOUSE STANDARD: PRT: 822 AA.
 ID CAD3_MOUSE STANDARD: PRT: 822 AA.
 AC P10287; 061465:

DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE CADHERIN-3 PRECURSOR (PLACENTAL-CADHERIN) (P-CADHERIN).
 CN CDH3 OR CDHF.
 GN Mus musculus (Mouse).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN 111

RP MEDLINE=88111554; PubMed=3428270;
 RA Nose A., Nagafuchi A., Takeichi M.;
 RT Isolation of placental cadherin cDNA: identification of a novel gene
 RT family of cell-cell adhesion molecules.";
 RL EMBO J. 6:3655-3661(1987).

CC PARTIAL SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=embryo;
 RX MEDLINE=91360343; PubMed=1866768;
 RA Hata M., Miyatani S., Copeland N.G., Gilbert D.J., Jenkins N.A.,
 RA Takeichi M.;
 RT "Genomic organization and chromosomal mapping of the mouse P-cadherin
 RT gene.";
 RL Nucleic Acids Res. 19:4437-4441(1991).
 RN 131

RP SEQUENCE OF 1-55 FROM N.A.
 RC STRAIN-C57BL6:
 RA Hattai M., Takeichi M.:
 RT "Complex cell type-specific transcriptional regulation by the promoter
 RL and an intron of the mouse p-cadherin gene";
 RN dev. Growth Differ. 36:509-519(1994).
 [4]
 RP DEVELOPMENTAL STAGE:
 RC STRAIN-C57BL6: TISSUE-Testis;
 RA MEDLINE-97033837; PubMed-8879495;
 RX Munro S.B., Blaschuk O.W.:
 RT "A comprehensive survey of the cadherins expressed in the testes of
 RL fetal, immature, and adult mice utilizing the polymerase chain
 RT reaction";
 RL Biol. Reprod. 55:822-827(1996).
 CC -I- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
 CC SORTING OF HETEROGENEOUS CELL TYPES.
 CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -I- DEVELOPMENTAL STAGE: EXPRESSION IS HIGH IN BOTH FETAL AND NEWBORN
 CC TESTIS BUT MINIMAL IN TESTIS OF 7-DAY-OLD ANIMALS. NOT DETECTED IN
 CC TESTIS OF 21-DAY-OLD OR ADULT.
 CC -I- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X06340; CAA29646.1; -
 CC EMBL: D12688; BAA02186.1; -
 CC PIR: S03163; IJMSCP.
 CC HSSP: P09803; ISUH.
 CC MGD: MGI:88356; Cdh3.
 CC InterPro: IPR002126; Cadherin.
 CC InterPro: IPR000233; Cadherin_C-term.
 CC Pfam: PF00028; cadherin_5.
 CC Pfam: PF01049; Cadherin_C-term; 1.
 CC PRINTS: PR00205; CADHERIN.
 CC SMART: SM00112; CA; 4.
 CC PROSITE: PS00232; CADHERIN_1; 3.
 CC PROSITE: PS50268; CADHERIN_2; 4.
 CC Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
 KW Signal.
 FT SIGNAL. 1 25 POTENTIAL.
 FT PROPEP 26 99
 FT CHAIN 100 822 CADHERIN-3.
 FT DOMAIN 100 647 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 648 670 POTENTIAL.
 FT DOMAIN 671 822 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 100 207 CADHERIN 1.
 FT DOMAIN 208 320 CADHERIN 2.
 FT DOMAIN 321 432 CADHERIN 3.
 FT DOMAIN 433 538 CADHERIN 4.
 FT DOMAIN 539 645 CADHERIN 5.
 FT DOMAIN 778 793 SER-RICH.
 FT CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 558 558 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT SEQUENCE 822 AA: 90754 MW: 8497502115B14DC3 CRC64;

Query Match 82.9%; Score 34; DB 1; Length 822;
 Best Local Similarity 80.0%; Pred. No. 1.4e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 WPXMR 6
 1 1 1
 DB 351 WPMR 355

RESULT 11
 ID RB52_CHRV1 STANDARD; PRT; 113 AA.
 AC P22860;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL CHAIN 2 (EC 4.1.1.39) (RUBISCO
 DE SMALL SUBUNIT 2).
 CN CBB52 OR RB5C.
 OS Chromatium vinosum.
 OC Bacteria; Proteobacteria; gamma subdivision; Chromatiales;
 OC Allochrocatium.
 CC NCBI_TaxID=1049;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC STRAIN-D / ATCC 17899 / DSM 180;
 CC MEDLINE-91138986; PubMed-1899846;
 CC Kobayashi H., Vale A.M., Takebe T., Akazawa T., Wada K.,
 CC Shiozaki K., Kobayashi K., Sugura M.:
 CC "Sequence and expression of genes encoding the large and small
 CC subunits of ribulose 1,5-bisphosphate carboxylase/oxygenase from
 CC Chromatium vinosum";
 CC Gene 97:55-62(1991).
 CC -I- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
 CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH
 CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
 CC ACTIVE SITE.
 CC -I- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) =
 CC 2 3-PHOSPHO-D-GLYCERATE.
 CC -I- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) =
 CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
 CC -I- SUBUNIT: 8 LARGE CHAINS + 8 SMALL CHAINS.
 CC -I- SIMILARITY: BELONGS TO THE RUBISCO SMALL CHAIN FAMILY.
 CC -I- CAUTION: IN C.VINOSUM TWO SIMILAR SET OF GENES CODE FOR RUBISCO
 CC LARGE AND SMALL CHAINS: THE RBCL-RBES AND THE RBCL-RBES SETS.
 CC UNDER STANDARD PHOTOAUTOTROPHIC CULTURE CONDITIONS ONLY THE LATER
 CC SET SEEMS ACTIVE, THE FORMER BEING PROBABLY CRYPTIC.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D90204; BAA14230.1; -
 CC PIR: J00587; RKR82.
 CC HSSP: P00870; IBUR.
 CC InterPro: IPR000894; Rubisco_small.
 CC Pfam: PF00101; Rubisco_small; 1.
 CC Pfam: PF000290; Rubisco_small; 1.
 CC Prodom: P000290; Rubisco_small; 1.
 CC Photosynthesis; Carbon dioxide fixation; Photorespiration;
 CC Lyase; Oxidoreductase; Monooxygenase; Multigene family.
 KW SEQUENCE 113 AA: 13391 MW: 002DDBCB86FEF399 CRC64;

Query Match 80.5%; Score 33; DB 1; Length 113;
 Best Local Similarity 50.0%; Pred. No. 34;
 Matches 3; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 FWPXMR 6
 : 1 1
 DB 58 WPMR 63

RESULT 12
 ID SOL4_YEAST STANDARD; PRT; 255 AA.
 SOL4_YEAST

AC p53315;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE PROBABLE 6-PHOSPHOGLUCONOLACTONASE 4 (EC 3.1.1.31) (6PGL).
 GN SOL4 OR YGR248W.
 OS *Saccharomyces cerevisiae* (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
 OX NCBI_TaxID=4932.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / FY1679;
 RX MEDLINE=97279234; PubMed=9133742;
 RA Ferrell F., Garigant G., Pavanello A., Guerreiro P., Azevedo D.,
 RA Rodriguez-Ponsaus C., Melchiorretto P., Panzeri L.,
 RA Agostoni Cardone M.L.;
 RT Analysis of a 17.9 kb region from *Saccharomyces cerevisiae*
 RT chromosome VII reveals the presence of eight open reading frames,
 RT including BRF1 (TFIRB70) and GCN5 genes.";
 RL Yeast 13:373-377(1997).
 CC -1- FUNCTION: HYDROLYSIS OF 6-PHOSPHOGLUCONOLACTONE TO 6-
 CC PHOSPHOGLUCONATE (BY SIMILARITY)
 CC -1- CATALYTIC ACTIVITY: 6-PHOSPHO-D-GLUCONO-1,5-LACTONE + H(2)O = 6-
 CC PHOSPHO-D-GLUCONATE.
 CC -1- PATHWAY: SECOND STEP IN PENTOSE PHOSPHATE PATHWAY.
 CC -1- SIMILARITY: BELONGS TO THE GLUCOSAMINE/GALACTOSAMINE-6-PHOSPHATE
 CC ISOMERASE FAMILY. 6-PHOSPHOGLUCONOLACTONASE SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: 273033; GAA97277.1; .
 DR SGD: S0003480; SOL4.
 UR InterPro: IPR000457; Glucosamine_iso.
 DR Pfam: PF01182; Glucosamine_iso.1.
 KM Hydrolyase.
 SO SEQUENCE 255 AA: 28447 MW: 691C9DDA04E5BEF6 CRC64;

Query Match 80.5% Score 33; DB 1; Length 255;
 Best Local Similarity 80.0% Pred No. 71;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 WPXMR 6
 11 11
 DB 66 WPMWR 70

RESULT 13
 ID PGSI_HUMAN STANDARD: PRT: 368 AA.
 AC P21810: P13247;
 DT 01-JAN-1996 (Rel. 13, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE BONE/CARTILAGE PROTEOGLYCAN I PRECURSOR (BIGLYCAN) (PG-S1).
 GN BGN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BONE;
 RX MEDLINE=89174714; PubMed=2647739;
 RX Fisher I.W., Termine J.D., Young M.F.;
 RT Deduced protein sequence of bone small proteoglycan I (biglycan)

RT shows homology with proteoglycan II (decorin) and several
 RT nonconnective tissue proteins in a variety of species.";
 RL J. Biol. Chem. 264:4571-4576(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91317791; PubMed=1860845;
 RA Fisher L.W., Heegaard A.M., Vetter U., Vogel W., Just W.,
 RA Termine J.D., Young M.F.;
 RT "Human biglycan gene. Putative promoter, intron-exon junctions, and
 RT chromosomal localization.";
 RL J. Biol. Chem. 266:14371-14377(1991).
 RN [3]
 RP SEQUENCE OF 38-57.
 RX MEDLINE=90073579; PubMed=2590169;
 RA Roughley P.J., White R.J.;
 RT "Dermatan sulphate proteoglycans of human articular cartilage. The
 RT properties of dermatan sulphate proteoglycans I and II.";
 RL Biochem. J. 262:823-827(1989).
 RN [4]
 RP SEQUENCE OF 38-66.
 RX MEDLINE=87250639; PubMed=3597437;
 RA Fisher L.W., Hawkins G.R., Turcos N., Termine J.D.;
 RT "Purification and partial characterization of small proteoglycans I
 RT and II, bone sialoproteins I and II, and osteonectin from the mineral
 RT compartment of developing human bone.";
 RL J. Biol. Chem. 262:9702-9708(1987).
 CC -1- TISSUE SPECIFICITY: FOUND IN THE EXTRACELLULAR MATRICES OF SEVERAL
 CC CONNECTIVE TISSUES, SPECIALLY IN ARTICULAR CARTILAGES.
 CC -1- PFM: THE TWO GLYCOSAMINOGLYCAN CHAINS ATTACHED TO BIGLYCAN CAN BE
 CC EITHER CHONDROITIN SULFATE OR DERMATAN SULFATE.
 CC -1- SIMILARITY: BELONGS TO THE SMALL INTERSTITIAL PROTEOGLYCANS
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: J04599; AAA36009.1; .
 DR EMBL: M65153; AAA52287.1; ALT_SEQ.
 DR EMBL: M65152; AAA52287.1; JOINED.
 DR PIR: A28457; A28457.
 DR PIR: A32458; A32458.
 DR PIR: A40757; A40757.
 DR PIR: S05639; S05639.
 DR MIM: 301870; .
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR000372; LRR_Nterm.
 DR InterPro: IPR003592; LRR_out.
 DR InterPro: IPR003591; LRR_typ.
 DR Pfam: PF00560; LRR_9.
 DR Pfam: PF01462; LRRNT.1.
 DR SMART: SM00370; LRR.3.
 DR SMART: SM00013; LRRNT.1.
 DR SMART: SM00369; LRR_TYP.1.
 KW Glycoprotein; Connective tissue; Extracellular matrix; Proteoglycan;
 KW Repeat; Leucine-rich repeat; Signal.
 FT SIGNAL 1 19
 FT PROPEP 20 37
 FT CHAIN 38 368 BONE/CARTILAGE PROTEOGLYCAN I.
 FT REPEAT 95 115 LRR 1.
 FT REPEAT 116 139 LRR 2.
 FT REPEAT 140 163 LRR 3.
 FT REPEAT 164 184 LRR 4.
 FT REPEAT 185 210 LRR 5.
 FT REPEAT 211 230 LRR 6.
 FT REPEAT 231 254 LRR 7.
 FT REPEAT 255 275 LRR 8.
 FT REPEAT 276 301 LRR 9.

FT REPEAT 302 320 LRR 10.
 FT CAROHD 42 42 O-LINKED (GLYCOSAMINOGLYCAN).
 FT CAROHD 47 47 O-LINKED (GLYCOSAMINOGLYCAN).
 FT CAROHD 270 270 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CAROHD 311 311 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DISULFID 63 76 BY SIMILARITY.
 FT DISULFID 321 354 BY SIMILARITY.
 FT DISULFID 139 140 KL -> NV (IN REF. 1).
 FT CONFLICT 163 164 EL -> DV (IN REF. 1).
 FT SEQUENCE 368 AA: 41654 MW: BFI6F304C5CDB3E CRC64:

Query Match: 80.5% Score 33: DB 1: Length 368:
 Best Local Similarity 80.0% Pred. No. 99:
 Matches 4: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

QY 2 WPXMR 6
 11 11
 DB 2 WPLMR 6

RESULT 14
 NASF_BACSU STANDARD: PRT: 483 AA.

AC P42437:
 ID NASF_BACSU STANDARD: PRT: 483 AA.
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE UROPORPHYRIN-III C-METHYLTRANSFERASE (EC 2.1.1.107) (UROGEN III
 METHYLASE) (SUMT) (UROPORPHYRINOGEN III METHYLASE) (UROM).
 GN NASF OR NASBE.
 OS Bacillus subtilis.
 OC Bacillus/Firmicutes: Bacillus/Clostridium group:
 OC Bacillus/Staphylococcus group: Bacillus.
 OX NCBI_TaxId=1423:
 RN 11
 RP SEQUENCE FROM N.A.

RC STRAIN-168:
 RX MEDLINE-95173124: PubMed=7868621:
 RA Ogawa K.-I., Akagawa E., Yamane K., Sun Z.-W., Lacelle M., Zuber P.,
 RA Nakano M.M.:
 RT "The nasb operon and nasa gene are required for nitrate/nitrite
 RT assimilation in Bacillus subtilis."
 RL J. Bacteriol. 177:1409-1413(1995).
 RN 121
 RP SEQUENCE FROM N.A.

RC STRAIN-168:
 RX MEDLINE-97124189: PubMed=8969502:
 RA Yamane K., Kumano M., Kurita K.:
 RT "The 25 degrees-36 degrees region of the Bacillus subtilis
 RT chromosome: determination of the sequence of a 146 kb segment and
 RT identification of 113 genes."
 RL Microbiology 142:3047-3056(1996).
 CC -1- CATALYTIC ACTIVITY: 2 S-ADENOSYL-L-METHIONINE + UROPORPHYRIN III
 CC -2 S-ADENOSYL-L-HOMOCYSTEINE + SIROHYDROCHLORIN.
 CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS SUMT, CYSO, CBIF/COBM
 CC AND CBIL/COBI.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: D30689: BAA06356.1: -
 DR EMBL: D50453: BAA08962.1: -
 DR EMBL: 299105: CAB12122.1: -
 DR Subtilisin: BGI1098: nasf.
 DR InterPro: IPR000878: Cortin_porph_methyltrnf.
 DR InterPro: IPR003043: Uropor_methyltrnsf.
 DR Pfam: PF00590: TP_methylase: 1.

DR PROSITE: PS00839: SUMT_1: 1.
 DR PROSITE: PS00840: SUMT_2: 1.
 KM Porphyrin biosynthesis: Transferase: Methyltransferase;
 KM Complete proteome.
 SQ SEQUENCE 483 AA: 53869 MW: 49752620FA96B74C CRC64:

Query Match: 80.5% Score 33: DB 1: Length 483:
 Best Local Similarity 80.0% Pred. No. 13e+02:
 Matches 4: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

QY 2 WPXMR 6
 11 11
 DB 284 WPXMR 288

RESULT 15
 MYLF_HUMAN STANDARD: PRT: 782 AA.

AC 000160:
 ID MYLF_HUMAN STANDARD: PRT: 782 AA.
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE MYOSIN IF (MYOSIN-IF) (FRAGMENT).
 GN MYOLF.
 OS Homo sapiens (Human).
 OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
 OC Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
 OX NCBI_TaxId=9606:
 RN 11
 RP SEQUENCE FROM N.A.

RC TISSUE=Retina:
 RX MEDLINE-97237053: PubMed=9119401:
 RA Crozet F., El Amraoui A., Blanchard S., Lenoir M., Ripoll C., Vago P.,
 RA Hamel C., Fitzames C., Levi-Acobas F., Depierre D., Mattei M.-G.,
 RA Weil D., Pujol R., Petit C.:
 RT "Cloning of the genes encoding two murine and human cochlear
 RT nonconventional type I myosins."
 RL Genomics 40:332-341(1997).
 CC -1- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE
 CC ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.
 CC THEIR HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANOUS
 CC COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS (BY
 CC SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE CLASS I FAMILY OF MYOSINS.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: X98411: CAAG7058.1: -
 DR HSSP: P08799: IMWG.
 DR InterPro: IPR000048: IQ.
 DR InterPro: IPR001452: SH3.
 DR InterPro: IPR001609: myosin_head.
 DR Pfam: PF00063: myosin_head: 1.
 DR Pfam: PF00018: SH3: 1.
 DR PRINTS: PR00193: MYOSINHEAVY.
 DR PRINTS: PR00452: SH3DOMAIN.
 DR PRODOM: PD000355: myosin_head: 1.
 DR SMART: SMO0015: IQ: 1.
 DR SMART: SMO0326: SH3: 1.
 DR PROSITE: PS50096: IQ: 1.
 DR PROSITE: PS50002: SH3: 1.
 DR Myosin: ATP-binding; Actin-binding; Calmodulin-binding; SH3 domain;
 KM Multigene family.
 FT NON_TER 1 1

DOMAIN <1 360 MYOSIN HEAD-LIKE.
-T DOMAIN 380 405 IQ.
FT DOMAIN 725 782 SH3.
FT NP_BIND 110 117 ATP (POTENTIAL).
FT DOMAIN 262 272 ACTIN-BINDING (POTENTIAL).
SQ SEQUENCE 782 AA: 89423 MW: 61EA9FB51E717B50 CRC64:

Query Match 80.5%: Score 33: DB 1: Length 782:
Best Local Similarity 80.0%: Pred. No. 1.9e+02:
Matches 4: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

QY 2 WPXWR 6
11 11
Db 325 WPRWR 329

Search completed: February 27, 2002, 11:42:50
Job time: 549 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:50:14 ; Search time 281.76 Seconds
(without alignments)
3.115 Million cell updates/sec

Title: US-09-446-109A-19
Perfect score: 41
Sequence: 1 FWPXMR 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 14627329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_17:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_mammal:*
- 6: sp_invertebrate:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_protent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	97.6	318	2	Q9KVD3
2	39	95.1	283	10	Q9LYX4
3	37	90.2	288	11	Q9CYC4
4	36	87.8	386	1	Q9HLY8
5	35	85.4	387	2	Q9I247
6	35	85.4	476	2	Q86506
7	35	85.4	605	5	Q10916
8	35	85.4	1612	5	Q9VYQ2
9	35	85.4	3164	12	Q9YTU2
10	35	85.4	3165	12	Q04350
11	35	85.4	3291	12	Q66225
12	34	82.9	89	6	Q28370
13	34	82.9	100	2	Q9S6V7
14	34	82.9	113	9	Q48445
15	34	82.9	129	4	Q9B858
16	34	82.9	148	1	Q9YFT5
17	34	82.9	159	4	Q9NUM6
18	34	82.9	173	2	Q9ACT2
19	34	82.9	176	10	Q9ARQ6

20	34	82.9	180	2	Q9KXL5	Q9KXL5 streptomyc
21	34	82.9	184	2	Q9ACW6	Q9ACW6 streptomyc
22	34	82.9	270	2	Q52690	Q52690 rhodobacter
23	34	82.9	288	6	Q9BE07	Q9BE07 macaca fasc
24	34	82.9	291	11	Q9CR58	Q9CR58 mus musculu
25	34	82.9	301	2	Q9I5M4	Q9I5M4 pseudomonas
26	34	82.9	307	5	Q9VVB3	Q9VVB3 drosophila
27	34	82.9	313	2	Q9RA26	Q9RA26 vibrio marl
28	34	82.9	315	11	Q9D7B5	Q9D7B5 mus musculu
29	34	82.9	322	4	Q9HC60	Q9HC60 homo sapien
30	34	82.9	322	11	Q9JMH0	Q9JMH0 ratu mus norv
31	34	82.9	325	11	Q9ESI8	Q9ESI8 mus musculu
32	34	82.9	325	11	Q9EP88	Q9EP88 ratu mus norv
33	34	82.9	350	8	Q9TKX0	Q9TKX0 nephroselm
34	34	82.9	353	4	Q9HC61	Q9HC61 homo sapien
35	34	82.9	373	1	Q58089	Q58089 pyrococcus
36	34	82.9	379	2	Q9EZ24	Q9EZ24 salmonella
37	34	82.9	401	5	Q9NKS1	Q9NKS1 leishmania
38	34	82.9	409	1	Q9UY92	Q9UY92 pyrococcus
39	34	82.9	423	5	Q20837	Q20837 caenorhabdl
40	34	82.9	501	10	Q9AXA9	Q9AXA9 oryza sativ
41	34	82.9	506	4	Q9BS52	Q9BS52 homo sapien
42	34	82.9	539	8	Q36792	Q36792 spinacia ol
43	34	82.9	539	10	Q41376	Q41376 spinacia ol
44	34	82.9	552	9	Q38064	Q38064 bacterioph
45	34	82.9	580	2	Q52803	Q52803 amycolatops

ALIGNMENTS

RESULT 1

Q9KVD3 1 PRELIMINARY: PRT: 318 AA.

AC Q9KVD3: 09KVD3:

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)

DE LIPID A BIOSYNTHESIS LAUROYL ACYLTRANSFERASE.

GN VC0213.

OS Vibrio cholerae.

OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.

OX NCBI_TaxID=666;

RA Fraser C.M.,

RP 11

RC STRAIN-EL TOR N16961 / SEROTYPE O1;

RC MEDLINE=20406833; PubMed=10952301;

RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,

RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,

RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,

RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragol I., Sellers P.,

RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,

RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,

RA Fraser C.M.,

KT "DNA sequence of both chromosomes of the cholera pathogen Vibrio

KT cholerae";

RL Nature 406:477-483(2000).

TR EMBL: AE004111; AAF93389.1; -

DR TIGR: VC0213; -

KW Transferase; Acyltransferase; Complete proteome.

SQ SEQUENCE 318 AA: 36342 MW: FE95D7A4C83106E1 CAC64:

Query Match 97.6%; Score 40; DB 2; Length 318;
Best Local Similarity 83.3%; Pred. NO. 35;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FWPXMR 6
111 11

DB 106 FWPXMR 111
RESULT 2

```

O9LYX4
ID O9LYX4 PRELIMINARY: PRT: 283 AA.
AC O9LYX4:
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE HYPOTHELICAL 32.7 KDA PROTEIN (GENOMIC DNA, CHROMOSOME 5, P1
CLONE:MOK16).
CN F15A17_140.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae:
OC eurosids.II: Brassicales: Brassicaceae: Arabidopsis.
OX NCBI_TaxID=3702;
RN 11
RP SEQUENCE FROM N.A.
RA Bevan M., Terry N., Ardiles W., Buysshart C., Dasseville R.,
RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,
RA Villarroel R., Gielens J., Van Montagu M., Bancroft I., Mewes H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (Apr-2000) to the EMBL/Genbank/DBJ databases.
RN 12
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (Apr-2000) to the EMBL/Genbank/DBJ databases.
RN 13
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA:
RX MEDLINE-97471969; PubMed-9330910;
RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
RA Miyajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
RT features of the 1.6 Mb regions covered by twenty physically assigned
RT clones."
RL DNA Res. 4:215-220(1997).
DR EMBL: AL163002; CAB86078.1; -
DR EMBL: AB005240; BAB08372.1; -
KW Hypothetical protein.
SQ SEQUENCE 283 AA: 32690 MW: CA01EAC858B997DC CRC64:

Query Match 95.1%; Score 39; DB 10; Length 283;
Best Local Similarity 83.3%; Pred. No. 44;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FWPXMR 6
DB 119 FWPXMR 124

RESULT 3
O9CYG4
ID O9CYG4 PRELIMINARY: PRT: 288 AA.
AC O9CYG4:
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE 5730478M09RIK PROTEIN.
CN 5730478M09RIK.
OS Mus musculus (Mouse).
OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
OC Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
OX NCBI_TaxID=10090;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE-21085660; PubMed-11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aikawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamahata I.,
RA Saito T., Okazaki Y., Ojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischiemann W., Gaasterland T., Gissi C., King B., Kochiwa H.,

```

```

RA Kuehl P., Lewis S., Matsuo Y., Mikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Steubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamlya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmink L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK017705; BAB30885.1; -
DR MGD: MGI:1917823; 5730478M09RIK.
SQ SEQUENCE 288 AA: 33042 MW: BD01C86D560835DA CRC64:

Query Match 90.2%; Score 37; DB 11; Length 288;
Best Local Similarity 66.7%; Pred. No. 91;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Caps 0;

OY 1 FWPXMR 6
DB 53 WPTWR 58

RESULT 4
O9HLY8
ID O9HLY8 PRELIMINARY: PRT: 386 AA.
AC O9HLY8:
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE HYPOTHELICAL PROTEIN TA0085.
CN TA0085.
OS Thermoplasma acidophilum.
OC Archaea: Euryarchaeota: Thermoplasmatales: Thermoplasmaceae:
OC Thermoplasmata.
OX NCBI_TaxID=2303;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN-DSM 1728;
RX MEDLINE-20479972; PubMed-11029001;
RA Ruepp A., Gräml W., Santos-Martinez M.-L., Korte K.K., Volker C.,
RA Mewes H.-W., Frisman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum."
RL Nature 407:508-513(2000).
DR EMBL: AL445063; CAC11233.1; -
DR InterPro: IPR001354; MR_MLE.
KW Pfam: PF01188; MR_MLE; 1.
KW Hypothetical protein: Complete proteome.
SQ SEQUENCE 386 AA: 43395 MW: 2334629296B5AA09 CRC64:

Query Match 87.8%; Score 36; DB 1; Length 386;
Best Local Similarity 66.7%; Pred. No. 176+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Caps 0;

OY 1 FWPXMR 6
DB 320 FWPXMR 325

RESULT 5
O91247
ID O91247 PRELIMINARY: PRT: 387 AA.
AC O91247:
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)

```

DE PROBABLE MFS TRANSPORTER.
 CN PA2068.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision: Pseudomonadaceae;
 CC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PA01;
 RA MEDLINE=20437337; PubMed=10984043;
 RX Stever C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Coltery L., Tolentino E., Westbrock-Madman S., Yuan Y.,
 RA Brody L.E., Coulter S.N., Folger K.R., Kas A., Lartig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reiter J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT *Complete genome sequence of Pseudomonas aeruginosa PA01, an
 opportunistic pathogen.*
 RL Nature 406:959-964(2000).
 DR EMBL: AE004633; AAC05456.1; -;
 KW Complete Proteome.
 SO SEQUENCE 387 AA; 41899 MW; 40B5592D8732F6FE CRC64;

Query Match 85.4%: Score 35; DB 2; Length 387;
 Best Local Similarity 80.0%: Pred. NO. 2.4e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;
 QY 1 FWPXW 5
 1111
 Db 18 FWPXW 22

RESULT 6
 086506 PRELIMINARY; PRT: 476 AA.
 ID 086506
 AC 086506;
 DT 01-NOV-1998 (TReMBLrel. 08, Created)
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE PUTATIVE TRANSMEMBRANE TRANSPORT PROTEIN.
 GN SCIC2.05C.
 OS Streptomyces coelicolor.
 CC Bacteria; Filicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2).
 RA Murphy L., Harris D.;
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2).
 RA Parkhill J., Barrett B.G., Rajandream M.A.;
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2).
 RA Redenbach M., Kleser H.M., Denapite D., Eichner A., Cullum J.,
 RA Khashl H., Hopwood D.A.;
 RT "A set of ordered cosmid and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome."
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL: AL031124; CAI19972.1; -;
 DR InterPro: IPR001248; Transp_cyt_pur.
 DR Pfam: Pf02133; Transp_cyt_pur; 1.
 KW Transmembrane.
 SO SEQUENCE 476 AA; 48910 MW; 652CA27EB5FB5DE CRC64;

Query Match 85.4%: Score 35; DB 2; Length 476;

Best Local Similarity 80.0%: Pred. NO. 2.8e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;
 QY 1 FWPXW 5
 1111
 Db 202 FWPXW 206

RESULT 7
 010916 PRELIMINARY; PRT: 605 AA.
 ID 010916
 AC 010916;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE HYPOTHEETICAL 70.1 KDA PROTEIN B0252.2 IN CHROMOSOME II.
 GN B0252.2.
 OS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 CC Rhabditidae; Peioderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL NZ;
 RA Du Z.;
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U23453; AAC46756.1; -;
 DR WormPeP: B0252.2; CE02418.
 DR InterPro: IPR000004; SAPB.
 DR InterPro: IPR000934; Ser_thr_Phosphatase.
 DR SMART: SM00118; SAPB; 1.
 KW Hypothetical protein.
 SO SEQUENCE 605 AA; 70063 MW; 9128E1A9708C44AD CRC64;

Query Match 85.4%: Score 35; DB 5; Length 605;
 Best Local Similarity 80.0%: Pred. NO. 3.5e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;
 QY 1 FWPXW 5
 1111
 Db 333 FWPXW 337

RESULT 8
 09Y02 PRELIMINARY; PRT: 1612 AA.
 ID 09Y02
 AC 09Y02;
 DT 01-MAY-2000 (TReMBLrel. 13, Created)
 DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE CG2779 PROTEIN.
 GN CG2779.
 OS Drosophila melanogaster (fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RA MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.C., Wortman J.K., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer V., Chapple M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunlov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jajall M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matzel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nussken D.R., Pacleb J.M.,
 RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Wotley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:
 RA "The genome sequence of *Drosophila melanogaster*."
 RA Science 287:2185-2195(2000).
 RL EMBL: AE003488; AAF48140.2; -
 DR Flybase: FBgn0030371; CG2779.
 DR InterPro: IPR002557; Chitin_binding.
 DR SMART: SM00494; ChBD2; 1.
 SO SEQUENCE 1612 AA; 149407 MW; D73877D13EF5F24C CRC64;

Query Match 85.4%; Score 35; DB 5; Length 1612;
 Best Local Similarity 66.7%; Pred. No. 8.4e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FWPXMR 6
 Db 1384 FWPXMR 1389

RESULT 9
 O9YTU2 PRELIMINARY; PRT: 3164 AA.
 AC O9YTU2;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE HYPOHETICAL 362.6 KDA PROTEIN.
 OS Cryphonectria hypovirus 1.
 OC Viruses; dsRNA viruses; Hypoviridae; Hypovirus.
 OX NCBI_TaxID=40281;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN=EURO7;
 RA Chen B., Nuss D.L.:
 RA Submitted (AUG-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF082191; AAD13750.1; -
 DR InterPro: IPR001410; DEAD.
 DR PROSITE: PS00589; PTS_HPR_SER; UNKNOWN_1.
 KW Hypothetical protein.

SO SEQUENCE 3164 AA; 362595 MW; 11C6BCA397855704 CRC64;

Query Match 85.4%; Score 35; DB 12; Length 3164;
 Best Local Similarity 66.7%; Pred. No. 1.5e+03;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FWPXMR 6
 Db 1002 FWPXMR 1007

RESULT 10
 O04350 PRELIMINARY; PRT: 3165 AA.
 AC O04350;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE HYPOHETICAL 361.7 KDA PROTEIN IN HYPOVIRULENCE-ASSOCIATED DS-RNA
 DE GENETIC ELEMENT.
 OS Cryphonectria hypovirus 1.
 OC Viruses; dsRNA viruses; Hypoviridae; Hypovirus.
 OX NCBI_TaxID=40281;
 RN 111
 RP SEQUENCE FROM N.A.
 RC MEDLINE=91184117; Pubmed=2009854;
 RA Shapira R., Choi G.H., Nuss D.L.:
 RA "Virus-like genetic organization and expression strategy for a double-
 XT stranded RNA genetic element associated with biological control of
 RL chestnut blight."
 RL EMO J. 10:731-739(1991).
 CC -
 CC WITH BIOLOGICAL CONTROL OF THE FUNGAL DISEASE CHESTNUT BLIGHT.
 CC THIS DS-RNA ARE ASSOCIATED WITH HYPOVIRULENCE. THEY ARE LOCALIZED
 CC IN THE CYTOPLASM.
 DR EMBL: M57938; AAA67458.1; -
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002114; PTS_HPR_ser.
 DR PROSITE: PS00589; PTS_HPR_SER; UNKNOWN_1.
 KW Hypothetical protein.
 SO SEQUENCE 3165 AA; 361715 MW; 732862977C2CD344 CRC64;

Query Match 85.4%; Score 35; DB 12; Length 3165;
 Best Local Similarity 66.7%; Pred. No. 1.5e+03;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FWPXMR 6
 Db 1002 FWPXMR 1007

RESULT 11
 O06225 PRELIMINARY; PRT: 3291 AA.
 AC O06225;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE ORFA AND ORFB, COMPLETE CDS.
 OS Cryphonectria hypovirus.
 OC Viruses; dsRNA viruses; Hypoviridae; Hypovirus.
 OX NCBI_TaxID=33772;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN=2;
 RA Hillman B.I., Halpern B.T., Brown M.P.:
 RA "A viral dsRNA element of the chestnut blight fungus with a distinct
 XT genetic organization."
 XT Virology 201:241-250(1994).
 KW EMBL: L29010; AAA20137.1; -

MEKOPS: C08.001: -
InterPro: IPR001410; DEAD.
DR InterPro: IPR002114; PTS_HPR_ser.
DR PROSITE: PS00589; PTS_HPR_SER; UNKNOWN_1.
SQ SEQUENCE 3291 AA: 377540 MW: A12A02D930D21239 CRC64:

Query Match 85.4%: Score 35; DB 12; Length 3291;
Best Local Similarity 66.7%: Pred. No. 1.6e+03;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FWPXW 6
DB 1041 FWPMQ 1046

RESULT 12
ID Q29370 PRELIMINARY: PRT: 89 AA.
AC Q29370:
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAY-1997 (TREMBlrel. 03, Last annotation update)
DE UNKNOWN PROTEIN (FRAGMENT).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE;
RX MEDLINE=96327607; Pubmed-8672129;
RA Winteroe A.K., Fredholm M., Davies W.;
RT "Evaluation and characterization of a porcine small intestine CDNA
RT library: analysis of 839 clones."
RL Mamm. Genome 7:509-517(1996).
DR EMBL: F15110; CAA23366.1; -;
FT NON_TER 1
FT NON_TER 89
SQ SEQUENCE 89 AA: 9699 MW: 09A6890EB371326C CRC64:

Query Match 82.9%: Score 34; DB 6; Length 89;
Best Local Similarity 80.0%: Pred. No. 91;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FWPXW 5
DB 85 FWPMW 89

RESULT 13
ID Q9S6V7 PRELIMINARY: PRT: 100 AA.
AC Q9S6V7:
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE FATTY ACID SYNTHASE (FRAGMENT).
OS Mycobacterium bovis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BCG-PASTEUR;
RA Fernandes N.D., Kolatukudy P.E.;
RT "Putative ketoacyl reductase of Mycobacterium bovis BCG fatty acid
RT synthetase gene."
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: U25719; AAC98804.1; -;
FT NON_TER 1
FT NON_TER 100

SQ SEQUENCE 100 AA: 11273 MW: D7E48ECF14071740 CRC64:

Query Match 82.9%: Score 34; DB 2; Length 100;
Best Local Similarity 80.0%: Pred. No. 1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 WPMW 6
DB 18 WPMW 22

RESULT 14
ID Q48445 PRELIMINARY: PRT: 113 AA.
AC Q48445:
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)
DE COMPLETE NUCLEOTIDE SEQUENCE.
CS Bacteriophage SPPI.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;
OC Lambda phage group.
OX NCBI_TaxID=10724;
RN [1]
RP SEQUENCE FROM N.A.
RA Alonso J.C., Luder G., Stiege A.C., Chai S., Weise F., Trautner T.A.;
KL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
LR EMBL: X97918; CAA66592.1; -;
SQ SEQUENCE 113 AA: 12538 MW: EC8B464F26D0944 CRC64:

Query Match 82.9%: Score 34; DB 9; Length 113;
Best Local Similarity 80.0%: Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FWPXW 5
DB 23 FWPLM 27

RESULT 15
ID Q9BR58 PRELIMINARY: PRT: 129 AA.
AC Q9BR58:
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE UNKNOWN (PROTEIN FOR MGC:2817).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY ADENOCARCINOMA;
RA Strausberg R.;
RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC006515; AAH06515.1; -;
SQ SEQUENCE 129 AA: 14975 MW: 12DE6FED0DC83FDE CRC64:

Query Match 82.9%: Score 34; DB 4; Length 129;
Best Local Similarity 80.0%: Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FWPXW 5
DB 50 FWPEM 54

Wed Feb 27 12:14:00 2002

us-09-446-109a-19.rspt

Page 6

Job time: 995 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:36:08 : Search time 132.19 Seconds
(without alignments)
1.021 Million cell updates/sec

Title: US-09-446-109A-19
Perfect score: 41
Sequence: 1 FMPXMR 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued_Patents_AA:*

- 1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
- 2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
- 3: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
- 4: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
- 5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:*
- 6: /cgn2_6/prodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	85.4	3165	2	US-08-459-146-3
2	35	85.4	3165	2	US-08-459-065-3
3	34	82.9	123	4	US-09-188-930-170
4	34	82.9	539	2	US-08-759-581B-20
5	34	82.9	822	2	US-08-474-067-9
6	34	82.9	822	2	US-08-474-068A-9
7	34	82.9	822	2	US-08-472-481-8
8	33	80.5	368	1	US-08-303-238-3
9	33	80.5	368	1	US-08-458-834-3
10	33	80.5	368	6	5340934-2
11	33	80.5	1312	4	US-09-041-886-19
12	32	78.0	12	4	US-08-702-054B-27
13	32	78.0	14	4	US-08-702-054B-4
14	32	78.0	406	3	US-08-934-494-2
15	32	78.0	406	3	US-09-143-068-2
16	30	73.2	18	4	US-08-702-054B-12
17	30	73.2	21	1	US-08-408-120-18
18	30	73.2	83	2	US-08-726-106A-76
19	30	73.2	117	1	US-07-614-443A-2
20	30	73.2	117	1	US-08-294-859-2
21	30	73.2	117	1	US-08-481-676-2
22	30	73.2	392	2	US-08-799-173A-7
23	30	73.2	568	1	US-07-862-021B-14
24	30	73.2	568	5	PCT-US93-03164-14
25	30	73.2	577	2	US-08-756-317-13
26	30	73.2	802	1	US-07-862-021B-12
27	30	73.2	802	1	US-08-313-288B-12

28	30	73.2	802	5	PCT-US93-03164-12	Sequence 12, Appl
29	30	73.2	807	1	US-07-862-021B-10	Sequence 10, Appl
30	30	73.2	807	1	US-08-313-288B-10	Sequence 10, Appl
31	30	73.2	807	5	PCT-US93-03164-10	Sequence 10, Appl
32	29.5	72.0	13	4	US-08-915-314-36	Sequence 36, Appl
33	29.5	72.0	12	4	US-08-915-314-81	Sequence 81, Appl
34	29	70.7	20	2	US-08-665-202-95	Sequence 95, Appl
35	29	70.7	66	2	US-08-751-767A-59	Sequence 59, Appl
36	29	70.7	71	3	US-08-801-092-8	Sequence 8, Appl
37	29	70.7	71	3	US-08-801-092-22	Sequence 22, Appl
38	29	70.7	71	3	US-08-801-092-36	Sequence 36, Appl
39	29	70.7	163	4	US-08-727-688-25	Sequence 25, Appl
40	29	70.7	209	4	US-09-164-193-8	Sequence 8, Appl
41	29	70.7	220	3	US-09-413-452-5	Sequence 5, Appl
42	29	70.7	220	4	US-09-413-452-1	Sequence 1, Appl
43	29	70.7	362	4	US-09-413-068-1	Sequence 1, Appl
44	29	70.7	362	4	US-09-413-068-1	Sequence 1, Appl
45	29	70.7	384	2	US-08-673-388-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-08-459-146-3
Sequence 3, Application US/08459146
Patent No. 5866405
GENERAL INFORMATION:
APPLICANT: Chol, Gil Ho
APPLICANT: Nuss, Donald Lee
TITLE OF INVENTION: Genetically Engineered Transmissible
TITLE OF INVENTION: Hypovirulence
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: George M. Gould, Esq., Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,146
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/832,117
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Roseman, Catherine R
REGISTRATION NUMBER: 34,240
REFERENCE/DOCKET NUMBER: 8589
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 235-6208
TELEFAX: (201) 235-3500
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3165 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHECAL: NO
ORIGINAL SOURCE:
ORGANISM: Endothia parasitica (Cryphonectria
parasitica)
STRAIN: EP713
US-08-459-146-3

Query Match 85.4%; Score 35; DB 2; Length 3165;
Best Local Similarity 66.7%; Pred. No. 4.5e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Caps 0;

OY 1 FWPXMR 6
DB 1002 FWPDMQ 1007

RESULT 2

US-08-459-065-3
Sequence 3; Application US/08459065

Patent No.: 5883642
GENERAL INFORMATION:

APPLICANT: Choi, Gil Ho
APPLICANT: Nuss, Donald Lee
TITLE OF INVENTION: Genetically Engineered Transmissible
TITLE OF INVENTION: Hypovirulence
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: George M. Gould, Esq., Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07110

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,065
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435

Prior Application Data:
APPLICATION NUMBER: US 07/832,117
FILING DATE: 06-FEB-1992

ATTORNEY/AGENT INFORMATION:
NAME: Roseman, Catherine R

REGISTRATION NUMBER: 34,240

REFERENCE/DOCKET NUMBER: 8589

TELECOMMUNICATION INFORMATION:

TELEPHONE: (201) 235-6208

TELEFAX: (201) 235-3500

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 3165 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ORIGINAL SOURCE:

ORGANISM: Endochloa parasitica (Cryphonectria

ORGANISM: parasitica)

STRAIN: EP713

US-08-459-065-3

Query Match 85.4%; Score 35; DB 2; Length 3165;
Best Local Similarity 66.7%; Pred. No. 4.5e+02;

Matches 4; Conservative 1; Mismatches 1; Indels 0; Caps 0;

OY 1 FWPXMR 6
DB 1002 FWPDMQ 1007

RESULT 3

US-09-188-930-170
Sequence 170; Application US/09188930A

Patent No. 6150502
GENERAL INFORMATION:

APPLICANT: Watson, James D.

APPLICANT: Strachan, Lorna

APPLICANT: Sleeman, Matthew

APPLICANT: Onrust, Rene

APPLICANT: Murison, James Greg

TITLE OF INVENTION: Compositions Isolated From Skin Cells

TITLE OF INVENTION: and Methods For Their Use

FILE REFERENCE: 11000.1011c1

CURRENT APPLICATION NUMBER: US/09/188,930A

CURRENT FILING DATE: 1998-11-09

NUMBER OF SEQ ID NOS: 348

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO: 170

LENGTH: 123

TYPE: PRT

ORGANISM: Rat

FEATURE:

NAME/KEY: UNSURE (27)

LOCATION: (27)...

NAME/KEY: UNSURE (104)

LOCATION: (104)...

NAME/KEY: UNSURE (118)

LOCATION: (118)...

US-09-188-930-170

Query Match 82.9%; Score 34; DB 4; Length 123;
Best Local Similarity 80.0%; Pred. No. 28;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 2 WPSMR 6
DB 14 WPSMR 18

RESULT 4

US-08-759-581B-20

Sequence 20; Application US/08759581B

Patent No. 5876945

GENERAL INFORMATION:

APPLICANT: CHISHOLM, DEXTER A.

APPLICANT: DINER, BRUCE A.

APPLICANT: DONALDSON, GAIL K.

APPLICANT: HERSHEY, HOWARD P.

APPLICANT: JORDAN, DOUGLAS B.

APPLICANT: TANG, XIAO-SONG

APPLICANT: TROST, JEFFREY T.

APPLICANT: WANG, SHAOJIE

APPLICANT: WARREN, PATRICK V.

TITLE OF INVENTION: METHODS FOR IDENTIFYING HERBICIDAL AGENTS THAT INHIBIT DI P

NUMBER OF SEQUENCES: 29

CORRESPONDENCE ADDRESS:

ADDRESSEE: E.I. DU PONT DE NEMOURS AND COMPANY

STREET: 1007 MARKET STREET

CITY: WILMINGTON

STATE: DELAWARE

COUNTRY: U.S.A.

ZIP: 19898

COMPUTER READABLE FORM:

MEDIUM TYPE: DISKETTE, 3.50 INCH

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: MICROSOFT WINDOWS 3.1

SOFTWARE: MICROSOFT WORD 2.0C

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/759,581B

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: FLOYD, LINDA AXAMETHY

REGISTRATION NUMBER: 33,692

REFERENCE/DOCKET NUMBER: CR-9964

TELECOMMUNICATION INFORMATION:

TELEPHONE: 302-892-8112

TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 539 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
ORIGINAL SOURCE:
STRAIN: SPINACH D1 PROTEASE PROTEIN
US-08-475-581B-20

Query Match 82.9%; Score 34; DB 2; Length 539;
Best Local Similarity 80.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 WPXMR 6
DB 79 WPSMR 83

RESULT 5
US-08-474-067-9
Sequence 9, Application US/08474067
Patent No. 5811518
GENERAL INFORMATION:
APPLICANT: Ranscht, Barbara
TITLE OF INVENTION: T-Cadherin Adhesion Molecule
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,067
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/213,361
FILING DATE: 14-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/607,293
FILING DATE: 30-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1682
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 822 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-474-067-9

Query Match 82.9%; Score 34; DB 2; Length 822;
Best Local Similarity 80.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 WPXMR 6
DB 11 11

DB 351 WPAWR 355

RESULT 6
US-08-474-068A-9
Sequence 9, Application US/08474068A
Patent No. 5837525
GENERAL INFORMATION:
APPLICANT: Ranscht, Barbara
TITLE OF INVENTION: T-Cadherin Adhesion Molecule
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,068A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/213,361
FILING DATE: 14-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/607,293
FILING DATE: 30-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1683
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 822 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-474-068A-9

Query Match 82.9%; Score 34; DB 2; Length 822;
Best Local Similarity 80.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 WPXMR 6
DB 351 WPAWR 355

RESULT 7
US-08-472-481-8
Sequence 8, Application US/08472481
Patent No. 5863804
GENERAL INFORMATION:
APPLICANT: Ranscht, Barbara
TITLE OF INVENTION: T-Cadherin Adhesion Molecule
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/472,481
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/213,361
FILING DATE: 14-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/607,293
FILING DATE: 30-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 1686
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 822 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-472-481-8

Query Match 82.9%; Score 34; DB 2; Length 822;
Best Local Similarity 80.0%; Pred. No. 1.8e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 WPXMR 6
1111
DB 351 WPMWR 355

RESULT 8
US-08-303-238-3
Sequence 3, Application US/08303238
Patent No. 5654270
GENERAL INFORMATION:
APPLICANT: RUOSLAHTI, ERKKI I.
APPLICANT: LONGAKER, MICHAEL T.
APPLICANT: WHITBY, DAVID J.
APPLICANT: HARPER, JOHN R.
APPLICANT: PIERSCHBACHER, MICHAEL D.
APPLICANT: BORDER, WAYNE A.
TITLE OF INVENTION: INHIBITORS OF CELL REGULATORY FACTORS
TITLE OF INVENTION: AND METHODS FOR PREVENTING OR REDUCING SCARRING
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: CAMPBELL AND FLORES
STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
CITY: SAN DIEGO
STATE: CALIFORNIA
COUNTRY: UNITED STATES
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/303,238
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/978,931
FILING DATE: 17-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: KONSKI, ANTOINETTE F.

REGISTRATION NUMBER: 34,202
REFERENCE/DOCKET NUMBER: P-LA 9453
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-535-9001
TELEFAX: 619-535-8949
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 368 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-303-238-3

Query Match 80.5%; Score 33; DB 1; Length 368;
Best Local Similarity 80.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 WPXMR 6
1111
DB 2 WPLMR 6

RESULT 9
US-08-458-834-3
Sequence 3, Application US/08458834
Patent No. 6277812
GENERAL INFORMATION:
APPLICANT: RUOSLAHTI, ERKKI I.
APPLICANT: LONGAKER, MICHAEL T.
APPLICANT: WHITBY, DAVID J.
APPLICANT: HARPER, JOHN R.
APPLICANT: PIERSCHBACHER, MICHAEL D.
APPLICANT: BORDER, WAYNE A.
TITLE OF INVENTION: INHIBITORS OF CELL REGULATORY FACTORS
TITLE OF INVENTION: AND METHODS FOR PREVENTING OR REDUCING SCARRING
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: CAMPBELL AND FLORES
STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700
CITY: SAN DIEGO
STATE: CALIFORNIA
COUNTRY: UNITED STATES
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,834
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/978,931
FILING DATE: 17-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: KONSKI, ANTOINETTE F.
REGISTRATION NUMBER: 34,202
REFERENCE/DOCKET NUMBER: P-LA 9453
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-535-9001
TELEFAX: 619-535-8949
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 368 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-458-834-3

Query Match 80.5%; Score 33; DB 4; Length 368;

- Local Similarity 80.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 WPXMR 6
1111
DB 2 WPLMR 6

RESULT 10
5340934-2
Patent No. 5340934
APPLICANT: TERMINE, JOHN D.; YOUNG, MARIAN F.; FISHER, LARRY W.
ROBEY, PAMELA G.
TITLE OF INVENTION: CDNA SEQUENCES OF HUMAN BONE MATRIX PROTEINS
NUMBER OF SEQUENCES: 13
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/432,044
FILING DATE: 03-NOV-1989
SEQ ID NO.: 2
LENGTH: 368
5340934-2

Query Match 80.5%; Score 33; DB 6; Length 368;
Best Local Similarity 80.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 WPXMR 6
1111
DB 2 WPLMR 6

RESULT 11
US-09-041-886-19
Sequence 19, Application US/09041886
Patent No. 6235872
GENERAL INFORMATION:
APPLICANT: Bredesen, Dale E.
TITLE OF INVENTION: Proapoptotic Peptides, Dependence
TITLE OF INVENTION: Polypeptides and Methods of Use
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESS: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/041,886
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 2626
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1312 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-041-886-19

Query Match 80.5%; Score 33; DB 4; Length 1312;
Best Local Similarity 80.0%; Pred. No. 4e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 WPXMR 6
1111
DB 25 WPCMR 29

RESULT 12
US-08-702-054B-27
Sequence 27, Application US/08702054B
Patent No. 6191254
GENERAL INFORMATION:
APPLICANT: Falls, Timothy J.
APPLICANT: Hancock, Robert E. W.
TITLE OF INVENTION: ANTIMICROBIAL CATIONIC PEPTIDES
TITLE OF INVENTION: AND METHODS OF SCREENING FOR THE SAME
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESS: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/702,054B
FILING DATE: 23-AUG-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/002,687
FILING DATE: 23-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: Haile, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07420/013001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-702-054B-27

Query Match 78.0%; Score 32; DB 4; Length 12;
Best Local Similarity 80.0%; Pred. No. 6.2;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 WPXMR 6
1111
DB 5 WPCMR 9

RESULT 13
US-08-702-054B-4
Sequence 4, Application US/08702054B
Patent No. 6191254
GENERAL INFORMATION:
APPLICANT: Falls, Timothy J.
APPLICANT: Hancock, Robert E. W.
TITLE OF INVENTION: ANTIMICROBIAL CATIONIC PEPTIDES

TITLE OF INVENTION: AND METHODS OF SCREENING FOR THE SAME
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/702,054B
FILING DATE: 23-AUG-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/002,687
FILING DATE: 23-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: Halle, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07420/013001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-702-054B-4

Query Match 78.0% Score 32: DB 4: Length 14:
Best Local Similarity 80.0% Pred. No. 7.3:
Matches 4: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

OY 2 WPXMR 6
11 11
DB 7 WPMWR 11

RESULT 14
US-08-934-494-2
Sequence 2, Application US/08934494
Patent No. 6030831
GENERAL INFORMATION:
APPLICANT: Gurney, Austin
TITLE OF INVENTION: Tie Ligands
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 MB floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/934,494
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: P1078

TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 406 amino acids
TYPE: Amino Acid
TOPOLOGY: linear
US-08-934-494-2

Query Match 78.0% Score 32: DB 3: Length 406:
Best Local Similarity 66.7% Pred. No. 1.9e+02:
Matches 4: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

OY 1 FWPXMR 6
11 11
DB 379 FWKTWR 384

RESULT 15
US-09-143-068-2
Sequence 2, Application US/09143068
Patent No. 6074873
GENERAL INFORMATION:
APPLICANT: Fong, Sherman
APPLICANT: Ferrara, Napoleone
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth
APPLICANT: Williams, Mickey
TITLE OF INVENTION: Tie Ligand Homologues
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 MB floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/143,068
FILING DATE: 28-Aug-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: P1078P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650/225-3216
TELEFAX: 650/952-9881
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 406 amino acids
TYPE: Amino Acid
TOPOLOGY: linear
US-09-143-068-2

Query Match 78.0% Score 32: DB 3: Length 406:
Best Local Similarity 66.7% Pred. No. 1.9e+02:
Matches 4: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

OY 1 FWPXMR 6
11 11
DB 379 FWKTWR 384

2:13:59 2002

us-09-446-109a-19.ra1

Page 7

Search completed: February 27, 2002, 11:36:09
Job time: 148 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:41:20 ; Search time 303.5 Seconds
(without alignments)

1.464 Million cell updates/sec

Title: US-09-446-109a-20

Perfect score: 33

Sequence: 1 FKXMR 6

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq.1101.*
1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID58/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID58/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID58/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	84.8	208	20	AAV60316 Human endometrium
2	28	84.8	1141	14	AAV31961 Human cardiac cgl
3	27	81.8	25	22	AAV15898 Peptide #2332 enco
4	27	81.8	25	22	AAV28405 Peptide #2442 enco
5	27	81.8	25	22	AAV03633 Peptide #2315 enco
6	27	81.8	117	22	AAV80944 Human nGPCR32 #1.
7	27	81.8	374	21	AAV74406 Neisseria meningit
8	27	81.8	376	21	AAV74404 Neisseria gonorrhoe
9	27	81.8	431	21	AAV71309 Human orphan G pro
10	27	81.8	431	21	AAV02843 Human G protein co
11	27	81.8	431	21	AAV02853 Human G protein co

12	27	81.8	431	21	AAV94993 Human secreted pro
13	27	81.8	431	22	AAV64289 Human GTP-binding
14	27	81.8	431	22	AAV74773 Human G protein-co
15	27	81.8	431	22	AAV48963 Human G protein-co
16	27	81.8	566	14	AAV37494 cdc25B. Homo sapi
17	27	81.8	566	18	AAV35317 Human cdc25B cell
18	27	81.8	566	19	AAV37987 Amino acid sequenc
19	27	81.8	566	19	AAV59136 Human cdc25B prote
20	27	81.8	566	21	AAV54906 Human cdc25B prote
21	27	81.8	566	22	AAV68002 Amino acid sequenc
22	27	81.8	566	22	AAV81085 Human Cdc25B prote
23	26	78.8	30	18	AAV10360 Hagfish intestinal
24	26	78.8	34	21	AAV07703 Arabidopsis thalia
25	26	78.8	35	21	AAV58719 Breast and ovarian
26	26	78.8	73	21	AAV56541 Human gastric can
27	26	78.8	98	22	AAV63505 Arabidopsis thalia
28	26	78.8	153	21	AAV24082 S. epidermidis ope
29	26	78.8	243	22	AAV82986 C glutamic prote
30	26	78.8	257	22	AAV92888 Arabidopsis thalia
31	26	78.8	281	21	AAV28556 Arabidopsis thalia
32	26	78.8	320	21	AAV28555 Arabidopsis thalia
33	26	78.8	347	21	AAV29090 Arabidopsis thalia
34	26	78.8	355	22	AAV83039 S. epidermidis ope
35	26	78.8	371	20	AAV29193 Arabidopsis thalia
36	26	78.8	397	21	AAV51955 Arabidopsis thalia
37	26	78.8	405	21	AAV44250 Arabidopsis thalia
38	26	78.8	405	21	AAV51954 Arabidopsis thalia
39	26	78.8	441	21	AAV29089 Arabidopsis thalia
40	26	78.8	480	21	AAV29088 Arabidopsis thalia
41	26	78.8	486	21	AAV51953 Arabidopsis thalia
42	26	78.8	499	21	AAV44249 Arabidopsis thalia
43	26	78.8	538	21	AAV41248 Arabidopsis thalia
44	26	78.8	578	21	AAV12991 MAV reverse transc
45	26	78.8	832	21	AAV12992 MAV reverse transc

ALIGNMENTS

RESULT 1

ID AAV60316 standard: Protein: 208 AA.

AAV60316:

31-JAN-2003 (first entry)

Human endometrium tumour EST encoded protein 376

Endometrium: human: tumour: cancer: anticancer: cytostatic: EST:

treatment: uterine: gene therapy: expressed sequence tag.

Homo sapiens

DE19817946

21-OCT-1998

17-APR-1998: 98DE-1017948.

17-APR-1998: 98DE-1017948.

(META-) METAGEN CES GENOMFORSCHUNG MBH.

Rosenthal A. Specht T. Hinzmann B. Schmitt A. Pilarsky C. Dahl E.

WPI: 1999-591957/51.

N-PSDB: AA742102.

New nucleic acid sequences expressed in uterine cancer tissues, and derived polypeptides for treatment of uterine and endometrial cancer and identification of therapeutic agents

PS Claim 23; Page 425; 444pp; German.

XX This invention describes novel human nucleic acid (cDNA) sequences (A),
 CC that are highly expressed in uterine tumour tissue and which have
 CC anticancer and cytostatic activity. (A) are used (i) for recombinant
 CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)
 CC are used (i) to identify agents suitable for treatment of uterine or
 CC endometrial cancer; (ii) directly for treating these forms of cancer
 CC (including expression from gene therapy vectors) and (iii) for
 CC generation of specific antibodies. (A) are identified by assembling ESTs
 CC (expressed sequence tags) from a particular tissue type before comparison
 CC of expression patterns. This allows a significantly longer fragment of
 CC the gene to be revealed, so should reduce the number of failures
 CC associated with the fact that ESTs from different libraries may represent
 CC different parts of the same unknown gene, distorting the estimated
 CC frequency of occurrence in a particular tissue. AAY59941-Y60328 represent
 CC protein fragments encoded by the human endometrium tumour cDNA library
 CC derived EST fragments represented in MA241981-242121.

XX SQ Sequence 208 AA:

Query Match 84.8%; Score 28; DB 20; Length 208;
 Best Local Similarity 66.7%; Pred. No. 2.4e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKMWR 6
 11: 11
 DB 7 fkvavr 12

RESULT 2
 AAR31961
 ID AAR31961 standard; Protein: 1141 AA.

XX AAR31961;
 AC
 XX 10-JUN-1993 (first entry)
 DT
 XX Human cardiac cgl PDE.
 DE
 XX CGMP: cAMP, phosphodiesterase; myocardial; positive inotropic agent;
 KW milrinone; amrinone; imazodan.
 KM
 XX Homo sapiens.
 OS
 XX US7801167-A.
 PN
 XX 15-DEC-1992.
 PD
 XX 03-DEC-1991; 91US-0801167.
 PF
 XX 03-DEC-1991; 91US-0801167.
 PR
 XX 03-DEC-1991; 91US-0801167.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
 PI Manganiello VC;
 XX WPI: 1993-067438/08.
 DR N-PSDB: AAO36780.
 XX
 XX New myocardial cAMP inhibited cAMP phosphodiesterase - useful as
 PT immunogens by being joined to polypeptide(s) e.g. haemocyanin,
 PT BSA, etc.
 PS Disclosure: Fig 1: 70pp; English.

XX Mixed oligodeoxynucleotide probes (P5 and P8) based on the partial
 CC amino acid sequences of two platelet phosphodiesterase peptides PDES
 CC and PDE8 were synthesised, as was an oligonucleotide based on a
 CC sequence from PDE 11A. The oligonucleotides were used to screen a
 CC human heart lambda ZAP II cDNA library and positive colonies
 CC purified by four successive screening. Clone n.13.2 contains the

CC entire cAMP- inhibited PDE cDNA coding sequence. Specific
 CC inhibition of myocardial cAMP inhibited cAMP phosphodiesterases is a
 CC primary mechanism of action for a number of positive inotropic
 CC agents, e.g. milrinone, amrinone, imazodan, etc. The protein may
 CC be bound to a solid support or joined to other polypeptides to be used
 CC as immunogens, e.g. keyhole limpet haemocyanin, BSA, tetanus toxoid etc.
 CC The protein (or fragments) may be used to raise monoclonal antibodies
 CC and hydromas. See also AAR31962.

XX SQ Sequence 1141 AA:

Query Match 84.8%; Score 28; DB 14; Length 1141;
 Best Local Similarity 66.7%; Pred. No. 1.2e+03;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKMWR 6
 11: 11
 DB 232 fkvavr 237

RESULT 3
 AAM15898
 ID AAM15898 standard; Protein: 25 AA.

XX AAM15898;
 AC
 XX 12-OCT-2001 (first entry)
 DT
 XX Peptide #2332 encoded by probe for measuring cervical gene expression.
 DE
 XX Probe: human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer.
 KM
 XX Homo sapiens.
 OS
 XX WO200157278-A2.
 PN
 XX 09-AUG-2001.
 PD
 XX 30-JAN-2001; 2001WO-US00670.
 PF
 XX 04-FEB-2000; 2000US-0180312.
 PR
 XX 26-MAY-2000; 2000US-0207456.
 PR
 XX 30-JUN-2000; 2000US-0608408.
 PR
 XX 03-AUG-2000; 2000US-0632366.
 PR
 XX 21-SEP-2000; 2000US-0234687.
 PR
 XX 27-SEP-2000; 2000US-0236359.
 PR
 XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI
 XX WPI: 2001-488901/53.
 DR
 XX
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human cervical epithelial cells -
 PT
 XX
 PS Claim 27; SEQ ID NO 20724; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes
 CC (SENP: see AAI10068-AI128459). The present sequence is a peptide encoded
 CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
 CC can be used to produce a single exon microarray, which can be used for
 CC measuring human gene expression in a sample derived from human cervical
 CC epithelial cells. By measuring gene expression, the probes are therefore
 CC useful in grading and/or staging of diseases of the cervix, notably
 CC cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 25 AA:

Query Match 81.8%; Score 27; DB 22; Length 25;
Best Local Similarity 66.7%; Pred. No. 46;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKMNR 6
11: 11
DB 5 fklmvr 10

RESULT 4
AAM28405
ID AAM28405 standard; Protein: 25 AA.

AC AAM28405;

DT 17-OCT-2001 (first entry)

DE Peptide #2442 encoded by probe for measuring placental gene expression.

XX Probe; microarray: human; placenta; antenatal diagnosis;

KW genetic disorder.

XX Homo sapiens.

OS MO200157272-A2.

XX 09-AUG-2001.

PD 30-JAN-2001; 2001WO-US000663.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PA Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI: 2001-488897/53.

DR Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human placenta -

PT Claim 27; SEQ ID NO 28674; 654bp; English.

XX The present invention relates to single exon nucleic acid probes (SENP;

CC see AAI31315-AA157546). The present sequence is a peptide encoded by one

CC such probe. The probes are useful for producing a microarray for

CC predicting, measuring and displaying gene expression in samples derived

CC from human placenta. The probes are useful for antenatal diagnosis of

CC human genetic disorders.

XX Sequence 25 AA:

Query Match 81.8%; Score 27; DB 22; Length 25;
Best Local Similarity 66.7%; Pred. No. 46;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKMNR 6
11: 11
DB 5 fklmvr 10

RESULT 5
AAM03633

ID AAM03633 standard; Protein: 25 AA.

AC AAM03633;

DT 09-OCT-2001 (first entry)

DE Peptide #2315 encoded by probe for measuring breast gene expression.

XX Probe; human; breast disease; breast cancer; development disorder;

KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX Homo sapiens.

OS MO200157270-A2.

XX 09-AUG-2001.

PD 29-JAN-2001; 2001WO-US000661.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PA Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI: 2001-476286/51.

DR Novel single exon nucleic acid probe used to measuring gene expression

XX in a human breast -

PT Claim 27; SEQ ID NO 12373; 322bp; English.

XX The present invention relates to novel single exon nucleic acid probes

CC (see AAI00010-AA110067). The present sequence is a peptide encoded by one

CC such probe. The probes are useful for measuring human gene expression in

CC a human breast sample, where the probe hybridises at high stringency to a

CC nucleic acid expressed in the human breast. The probes are useful for

CC predicting, diagnosing, grading, staging, monitoring and prognosing

CC diseases of the human breast, particularly those diseases with polygenic

CC aetiology. The diseases include: breast cancer, disorders of development,

CC inflammatory diseases of the breast, fibrocystic changes, proliferative

CC breast disease and non-carcinoma tumours.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 25 AA:

Query Match 81.8%; Score 27; DB 22; Length 25;
Best Local Similarity 66.7%; Pred. No. 46;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKMNR 6
11: 11
DB 5 fklmvr 10

RESULT 6
AAG80944
ID AAG80944 standard; Protein: 117 AA.

AC AAG80944;

DT 28-AUG-2001 (first entry)

XX Human nPCR32 #1.

XX G protein-coupled receptor; nGPCR; seven transmembrane receptor;
 KW signal transduction; schizophrenia; thyroid disorder; renal failure;
 KW rheumatoid arthritis; CNS disorder; infection; metabolic disease;
 KW cardiovascular disease; proliferative disorder; hormonal disorder;
 KW neurological disorder; neuronal disorder; Alzheimer's disease; cancer;
 KW attention deficit-hyperactivity disorder/attention deficit disorder;
 KW Parkinson's disease; migraine; senile dementia; inflammatory disease;
 KW rheumatoid arthritis; autoimmune disorder; respiratory ailment;
 KW neuroprotective;
 KW
 OS Homo sapiens.
 XX
 PN W0200136473-A2.
 PD
 XX 25-MAY-2001.
 XX
 PF 16-NOV-2000; 2000WO-US31581.
 XX
 PR 16-NOV-1999; 99US-0165838.
 PR 17-NOV-1999; 99US-0166071.
 PR 19-NOV-1999; 99US-0166678.
 PR 28-DEC-1999; 99US-0173396.
 PR 22-FEB-2000; 2000US-0184129.
 PR 28-FEB-2000; 2000US-0185421.
 PR 28-FEB-2000; 2000US-0185554.
 PR 02-MAR-2000; 2000US-0186530.
 PR 03-MAR-2000; 2000US-0186811.
 PR 09-MAR-2000; 2000US-0188114.
 PR 17-MAR-2000; 2000US-0190310.
 PR 21-MAR-2000; 2000US-0190800.
 PR 20-APR-2000; 2000US-0198568.
 PR 02-MAY-2000; 2000US-0201190.
 PR 08-MAY-2000; 2000US-0203111.
 PR 25-MAY-2000; 2000US-0207094.
 XX
 XX (PMAA) PHARMACIA & UPJOHN CO.
 PA
 XX Vogell G, Wood LS, Parodi LA, Hiebsch RR, Lind P, Slightom J;
 PI Schellin KA, Kaytes PS, Bannigan CM, Ruff V, Sejlitz T, Huff RM;
 XX
 DR WPI: 2001-389826/41.
 DR N-PSDB: AAH50984.
 XX
 PT New G protein-coupled receptor (nGPCR-x) and its encoding
 PT polynucleotide useful for diagnosing and treating e.g. schizophrenia .
 XX
 PS Claim 37: Page 80; 26pp; English.

Query Match 81.8%; Score 27; DB 22; Length 117;
 Best Local Similarity 66.7%; Pred. No. 2.1e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 FKMWR 6
 111 1:
 Db 83 fkmwq 88
 RESULT 7
 AA74406
 ID AA74406 standard; Protein: 374 AA.
 AC
 XX AA74406;
 XX
 DT 21-MAR-2000 (first entry)
 XX
 DE Neisseria meningitidis ORF 088 protein sequence SEQ ID NO:288.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KW antibacterial; gene therapy.
 XX
 OS Neisseria meningitidis.
 PN W09957280-A2.
 PD
 XX 11-NOV-1999.
 XX
 PF 30-APR-1999; 99WO-US09346.
 XX
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 XX Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;
 PI Petersen J, Piza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 DR WPI: 2000-062150/05.
 DR N-PSDB: AA253168.
 XX
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics .
 XX
 PS Claim 2: Page 282; 1453pp; English.

AA253015 to AA254536, AA254577 to AA254615, and AA74253 to AA75941
 represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX
 SO Sequence 374 AA;

Query Match 81.8%; Score 27; DB 21; Length 374;

Best Local Similarity 66.7%; Pred. No. 6.6e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FRMKXR 6
11111
Db 131 fkmwq 136

RESULT 8

AA774404
ID AA774404 standard; Protein: 376 AA.

XX AC AA774404;

XX DT 21-MAR-2000 (first entry)

XX DE Neisseria gonorrhoeae ORF 088 protein sequence SEQ ID NO:284.

KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;

KW antibacterial; gene therapy.

OS Neisseria gonorrhoeae.

PN MO9957280-A2.

XX PD 11-NOV-1999.

XX PF 30-APR-1999; 99WO-US09346.

XX PR 01-MAY-1998; 98US-0083758.

XX PR 31-JUL-1998; 98US-0094869.

XX PR 02-SEP-1998; 98US-0098994.

XX PR 09-OCT-1998; 98US-0103749.

XX PR 09-OCT-1998; 98US-0103794.

XX PR 25-FEB-1999; 99US-0121528.

XX PA (CHIR) CHIRON CORP.

XX PA (GENO-) INST GENOMIC RES.

PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;

XX DR MPI: 2000-062150/05.

XX DR N-PSDB: AA253166.

PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics

XX PS Claim 2: Page 280-281; 1453pp; English.

XX AA253015 to AA254536, AA254577 to AA254615, and AA774253 to AA775941
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia). To detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.

XX Sequence 376 AA;

Query Match 81.8%; Score 27; DB 21; Length 376;
Best Local Similarity 66.7%; Pred. No. 6.6e+02;

Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FRMKXR 6
11111
Db 131 fkmwq 136

RESULT 9

AA771309
ID AA771309 standard; Protein: 431 AA.

XX AC AA771309;

XX DT 02-NOV-2000 (first entry)

XX DE Human orphan G protein-coupled receptor hrup4.

KW Human; orphan G protein-coupled receptor; GPCR; hrup4; drug screening;
transmembrane receptor; expressed sequence tag; EST; signal cascade.

OS Homo sapiens.

PN WO200031258-A2.

XX PD 02-JUN-2000.

XX PF 13-OCT-1999; 99WO-US23687.

XX PR 20-NOV-1998; 98US-0109213.

XX PR 16-FEB-1999; 99US-0120416.

XX PR 26-FEB-1999; 99US-0121852.

XX PR 12-MAR-1999; 99US-0123946.

XX PR 28-MAY-1999; 99US-0136436.

XX PR 28-MAY-1999; 99US-0136437.

XX PR 28-MAY-1999; 99US-0136439.

XX PR 28-MAY-1999; 99US-0136567.

XX PR 28-MAY-1999; 99US-0137127.

XX PR 28-MAY-1999; 99US-0137131.

XX PR 29-JUN-1999; 99US-0141448.

XX PR 29-SEP-1999; 99US-0156555.

XX PR 29-SEP-1999; 99US-0156633.

XX PR 29-SEP-1999; 99US-0156634.

XX PR 29-SEP-1999; 99US-0156635.

XX PR 01-OCT-1999; 99US-0157280.

XX PR 01-OCT-1999; 99US-0157281.

XX PR 01-OCT-1999; 99US-0157282.

XX PR 01-OCT-1999; 99US-0157293.

XX PR 01-OCT-1999; 99US-0157294.

XX PR 12-OCT-1999; 99US-0416760.

XX PR 12-OCT-1999; 99US-0417044.

XX PA (AREN-) ARENA PHARM INC.

XX PI Chen R, Dang HT, Liao CW, Lin I;

XX DR MPI: 2000-400068/34.

XX DR N-PSDB: AAD01136.

PT Novel human orphan G protein-coupled receptors and the encoding cDNAs
PT for use in the identification of G protein-coupled receptor agonists -
PT Claim 74: Page 80-91; 102pp; English.

The present amino acid sequence is the hrup4, an endogenous human
CC orphan G protein-coupled receptor (GPCR). The full length hrup4 cDNA was
CC cloned by RT-PCR with human brain cDNA as template. The hrup4 PCR
CC fragment obtained was an alternatively spliced form of the EST (expressed
CC sequence tag) clone A1307658. The orphan GPCR of the invention, like
CC all GPCRs has seven transmembrane alpha helices with an extracellular
CC N-terminus and an intracellular C-terminus. However, no endogenous
CC ligands has yet been identified for the proteins of the invention. The
CC orphan GPCRs may be used in the identification of their endogenous

CC ligands, and to screen potential GPCR agonists and antagonists for use as
CC pharmaceutical agents. The proteins may also be used in the study of
CC GPCR-mediated signalling cascades, and to elucidate their precise role in
CC normal and diseased human conditions. Nucleic acid encoding human orphan
CC GPCRs may be used for tissue localisation expression analysis to provide
CC information about their function in healthy and pathological states.

XX Sequence 431 AA:

Query Match: 81.8%; Score 27; DB 21; Length 431;
Best Local Similarity 66.7%; Pred. No. 7.5e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKMWR 6
1111
DB 151 fkmwq 156

RESULT 10

AAB02843
ID AAB02843 standard; Protein: 431 AA.

XX AAB02843;

DT 22-AUG-2000 (first entry)

XX Human G protein coupled receptor hRUP4 protein SEQ ID NO:40.

DE Human: G protein coupled receptor; GPCR; transmembrane receptor;
KW identification; agonist; screening; therapeutic; pharmaceutical;
KW mutant.

OS Homo sapiens.

XX MO200022131-A2.

XX 20-APR-2000.

PF 13-OCT-1999; 99WO-US24065.

XX 13-OCT-1998; 98US-0170496.

PR 12-NOV-1998; 98US-0108029.

PR 20-NOV-1998; 98US-0109213.

PR 27-NOV-1998; 98US-0110060.

PR 16-FEB-1999; 99US-0120416.

PR 26-FEB-1999; 99US-0121852.

PR 12-MAR-1999; 99US-0123944.

PR 12-MAR-1999; 99US-0123945.

PR 12-MAR-1999; 99US-0123946.

PR 12-MAR-1999; 99US-0123949.

PR 12-MAR-1999; 99US-0123951.

PR 28-MAY-1999; 99US-0136436.

PR 28-MAY-1999; 99US-0136437.

PR 28-MAY-1999; 99US-0137127.

PR 28-MAY-1999; 99US-0137131.

PR 28-MAY-1999; 99US-0137567.

PR 30-JUN-1999; 99US-0141448.

PR 27-AUG-1999; 99US-0151114.

PR 03-SEP-1999; 99US-0152524.

PR 29-SEP-1999; 99US-0156633.

PR 29-SEP-1999; 99US-0156555.

PR 29-SEP-1999; 99US-0156634.

XX (AREN-) ARENA PHARM INC.

PA Behan DP, Lehmann-Brunsmma K, Chalmer DT, Chen R, Dang HT;

PI Core M, Liao CM, Lin I, Lowitz K, White C;

XX WPI: 2000-317986/27.

DR N-PSDB: AAA46037.

XX Non-endogenous, human G protein-coupled receptors for screening
PT receptor, inverse or partial agonists useful as therapeutic agents
XX Example 1: Page 119-120, 187pp: English.

XX The present invention describes transmembrane receptors, preferably
CC human G protein coupled receptors (GPCR), for which the endogenous
CC ligand is unknown (orphan GPCR receptors). More specifically the present
CC invention relates to non-endogenous, constitutively activated versions
CC of a human GPCR. These non-endogenous human GPCRs can be useful for
CC the direct identification of candidate compounds as receptors agonists,
CC inverse agonists or partial agonists for use as pharmaceutical agents.
CC AAA46017 to AAA46126 and AAB02825 to AAB02859 represent sequences used in
CC the exemplification of the present invention.

XX Sequence 431 AA:

Query Match: 81.8%; Score 27; DB 21; Length 431;
Best Local Similarity 66.7%; Pred. No. 7.5e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKMWR 6
1111
DB 151 fkmwq 156

RESULT 11

AAB02853
ID AAB02853 standard; Protein: 431 AA.

XX AAB02853;

DT 22-AUG-2000 (first entry)

XX Human G protein coupled receptor hRUP4 (V272K) protein SEQ ID NO:128.

DE Human: G protein coupled receptor; GPCR; transmembrane receptor;
KW identification; agonist; screening; therapeutic; pharmaceutical;
KW mutant.

OS Homo sapiens.

XX Synthetic.

XX MO200022131-A2.

XX 20-APR-2000.

PF 13-OCT-1999; 99WO-US24065.

XX 13-OCT-1998; 98US-0170496.

PR 12-NOV-1998; 98US-0108029.

PR 20-NOV-1998; 98US-0109213.

PR 27-NOV-1998; 98US-0110060.

PR 16-FEB-1999; 99US-0120416.

PR 26-FEB-1999; 99US-0121852.

PR 12-MAR-1999; 99US-0123944.

PR 12-MAR-1999; 99US-0123945.

PR 12-MAR-1999; 99US-0123946.

PR 12-MAR-1999; 99US-0123949.

PR 12-MAR-1999; 99US-0123951.

PR 28-MAY-1999; 99US-0136436.

PR 28-MAY-1999; 99US-0136437.

PR 28-MAY-1999; 99US-0137127.

PR 28-MAY-1999; 99US-0137131.

PR 28-MAY-1999; 99US-0137567.

PR 30-JUN-1999; 99US-0141448.

PR 27-AUG-1999; 99US-0151114.

PR 03-SEP-1999; 99US-0152524.

PR 29-SEP-1999; 99US-0156633.

PR 29-SEP-1999; 99US-0156555.
 PR 29-SEP-1999; 99US-0156634.
 XX
 PA (AREN-) ARENA PHARM INC.
 XX Behan DP, Lehmann-Brulisma K, Chalmers DT, Chen R, Dang HT;
 PI Gore M, Liaw CW, Lin I, Lowitz K, White C;
 XX
 DR WPI: 2000-317986/27.
 DR N-PSDB: AAA46115.
 PT Non-endogenous, human G protein-coupled receptors for screening
 PT receptor, inverse or partial agonists useful as therapeutic agents
 XX
 PS Example 2: Page 164-166; 187pp; English.
 XX
 CC The present invention describes transmembrane receptors, preferably
 CC human G protein coupled receptors (GPCR), for which the endogenous
 CC ligand is unknown (orphan GPCR receptors). More specifically the present
 CC invention relates to non-endogenous, constitutively activated versions
 CC of a human GPCR. These non-endogenous human GPCRs can be useful for
 CC the direct identification of candidate compounds as receptors agonists,
 CC inverse agonists or partial agonists for use as pharmaceutical agents.
 CC AAA46017 to AAA46126 and AAB02825 to AAB02859 represent sequences used in
 CC the exemplification of the present invention.
 XX
 SQ Sequence 431 AA:
 QY
 DB 151 fkmkqw 156
 QY 1 fkmkxr 6
 III I:
 DB 151 fkmkqw 156
 RESULT 12
 AAY94993
 ID AAY94993 standard; Protein: 431 AA.
 XX
 AC AAY94993;
 XX
 DT 19-JUN-2000 (first entry)
 XX
 DE Human secreted protein vc38_l, SEQ ID NO:26.
 XX
 KW Human; secreted protein; cancer; tumour; cardiovascular disorder;
 KW blood disorder; haemophilia; autoimmune disease; diabetes; inflammation;
 KW infection; fungal; bacterial; viral; HIV; allergy; arthritis;
 KW neurodegenerative disease; asthma; contraceptive.
 XX
 OS Homo sapiens.
 XX
 PN WO200011015-A1.
 XX
 PD 02-MAR-2000.
 PD
 PF 24-AUG-1999; 99WO-US19351.
 XX
 XX 24-AUG-1998; 98US-0097638.
 PR 24-AUG-1998; 98US-0097659.
 PR 09-SEP-1998; 98US-0099618.
 PR 28-SEP-1998; 98US-0102092.
 PR 25-NOV-1998; 98US-0109978.
 PR 23-DEC-1998; 98US-0113645.
 PR 23-DEC-1998; 98US-0113646.
 PR 23-AUG-1999; 99US-0379246.
 PR
 XX (ALPH-) ALPHAGENE INC.
 PA
 XX Valenzuela D, Yuan O, Hoffman H, Hall J, Rapiejko P;

XX
 DR WPI: 2000-224657/19.
 DR
 FT New secreted or transmembrane proteins and polynucleotides encoding
 FT them, useful for treating neurodegenerative disorders, autoimmune
 PT diseases and cancer -
 XX
 PS Claim 35; Page 284-285; 357pp; English.
 XX
 CC The invention relates to 40 human secreted proteins (AAY94981-Y95020),
 CC and cDNA sequences encoding them (AAA23423-A23462). The secreted
 CC proteins of the invention include those that are thought to be only
 CC partially secreted, i.e., transmembrane proteins. The proteins of the
 CC invention may exhibit one or more activities selected from the following:
 CC cytokine activity; cell proliferation; differentiation; immune
 CC modulation; haematopoiesis regulation; tissue growth activity;
 CC activin/inhibin activity; chemotactic/chemokinetic activity; haemostatic
 CC and thrombolytic activity; anti-inflammatory activity; and tumour
 CC inhibition activity. The proteins may be administered to patients as
 CC vaccines, and the nucleotides may be used as part of a gene therapy
 CC regime. Diseases or conditions that may be treated using the proteins or
 CC nucleotides of the invention include autoimmune diseases; genetic
 CC disorders; haemophilia; cardiovascular diseases; cancer; bacterial,
 CC fungal and viral infections, especially HIV; multiple sclerosis;
 CC rheumatoid arthritis; pulmonary inflammation; Guillain-Barre syndrome;
 CC insulin dependent diabetes mellitus; and allergic reactions such as
 CC asthma and anaemia. They may also be used for treating wounds, burns,
 CC ulcers, osteoporosis, osteoarthritis, periodontal diseases, Alzheimer's
 CC disease, Parkinson's disease, Huntington's disease and amyotrophic
 CC lateral sclerosis (ALS). Proteins with activin/inhibin activity may
 CC additionally be useful as contraceptives. Nucleic acid sequences of the
 CC invention may be used in chromosome mapping, and as a source of
 CC diagnostic primers and probes. The present sequence represents one of the
 CC 40 proteins of the invention.
 XX
 SQ Sequence 431 AA:
 QY
 DB 151 fkmkqw 156
 QY 1 fkmkxr 6
 III I:
 DB 151 fkmkqw 156
 RESULT 13
 AAG64289
 ID AAG64289 standard; Protein: 431 AA.
 XX
 AC AAG64289;
 XX
 DT 21-SEP-2001 (first entry)
 XX
 DE Human GTP-binding protein-coupled receptor GPRV11.
 XX
 KW GTP-binding protein-coupled receptor; neuroprotective; immunomodulatory;
 KW muscular; urinary; circulatory; anorectic; human; guanosine triphosphate;
 KW G-protein.
 XX
 OS Homo sapiens.
 XX
 PN WO200148189-A1.
 XX
 PD 05-JUL-2001.
 PD
 PF 28-DEC-2000; 2000WO-JP09409.
 XX
 XX 28-DEC-1999; 99JP-0375152.
 PR 31-MAR-2000; 2000JP-0101339.
 PR 23-MAY-2000; 2000JP-0155978.
 PR
 XX

PA (HELI-) HELIX RES INST.
 XX
 PI Matsumoto S, Oda T, Saito Y, Morikawa N, Yoshida K, Suwa M;
 PI Sugiyama T;
 DR WPI: 2001-425663/45.
 DR N-PSDB: AAF81819.
 XX
 PT Family of guanosine triphosphate binding protein coupled receptors and
 PT genes encoding them for treatment and prevention of diseases associated
 PT with these receptors
 XX
 PS Claim 1: Pages 59-62; 137pp; Japanese.
 XX
 CC The present sequence is the protein sequence for a human guanosine
 CC triphosphate (GTP)-binding protein-coupled receptor. The receptor is
 CC useful for the investigation, diagnosis, treatment and prevention of
 CC diseases associated with GTP-binding protein-coupled receptors, including
 CC neurological, circulatory, digestive system, immune system, muscle and
 CC urinary system disorders. GTP-binding proteins are also known as
 CC G-proteins.
 CC
 SO Sequence 431 AA:
 Query Match 81.8%; Score 27; DB 22; Length 431;
 Best Local Similarity 66.7%; Pred. No. 7.5e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FRMKWR 6
 ||| |:
 Db 151 fkmkqw 156
 RESULT 14
 AAB74773
 ID AAB74773 standard; Protein: 431 AA.
 XX
 AC AAB74773:
 XX
 DT 05-JUN-2001 (first entry)
 XX
 DE Human G protein-coupled receptor protein A027 SEQ ID NO:1.
 XX
 KW Human: G protein-coupled receptor; A027; nootropic; neuroprotective;
 KW hypotensive; antirheumatic; antiallergic; cardiant; antianginal;
 KW abortifacient; gene therapy; Alzheimer's disease; hypertension;
 KW pregnancy termination; rheumatism; allergy; angina pectoris.
 XX
 OS Homo sapiens.
 XX
 PN WO200116316-A1.
 XX
 PD 08-MAR-2001.
 XX
 PE 24-AUG-2000; 2000WO-JP05684.
 XX
 PR 27-AUG-1999; 99JP-0241530.
 XX
 PA (TAKE) TAKEDA CHEM IND LTD.
 XX
 PI Watanabe T, Kikuchi K, Shintani Y;
 XX
 DR WPI: 2001-226689/23.
 DR N-PSDB: AAF81819, AAF81819.
 XX
 PT Human brain-originated guanosine triphosphate protein-coupled receptor,
 PT its salt and encoded gene, useful in (gene) diagnosis and development
 PT of preventives and remedies for e.g. Alzheimer disease, hypertension
 PT and rheumatism
 XX
 NS Claim 1: Page 76-78; 84pp; Japanese.
 XX

CC The present sequence represents a human G protein-coupled receptor
 CC designated A027. A027 has nootropic, neuroprotective, hypotensive,
 CC antirheumatic, antiallergic, cardiant, antianginal and abortifacient
 CC activities. A027 can be used as a G protein-coupled receptor protein
 CC agonist, G-protein-coupled receptor signal transducer and in gene
 CC therapy. A027 can be sued in the diagnosis and development of
 CC preventives and remedies for diseases associated with dysfunction of
 CC A027 e.g. Alzheimer's disease, hypertension, pregnancy termination,
 CC rheumatism, allergy and angina pectoris.
 CC
 SO Sequence 431 AA:
 Query Match 81.8%; Score 27; DB 22; Length 431;
 Best Local Similarity 66.7%; Pred. No. 7.5e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FRMKWR 6
 ||| |:
 Db 151 fkmkqw 156
 RESULT 15
 AAB48963
 ID AAB48963 standard; Protein: 431 AA.
 XX
 AC AAB48963:
 XX
 DT 27-MAR-2001 (first entry)
 XX
 DE Human G protein-coupled receptor AXOR16.
 XX
 KW AXOR16: human: G protein-coupled receptor; 7TM receptor;
 KW chromosome 11q12.2; infection: viral; bacterial; fungal;
 KW protozoan; HIV-1; HIV-2; pain; cancer; diabetes; obesity;
 KW anorexia; bulimia; osteoporosis; asthma; allergy; urinary retention;
 KW acute heart failure; hypotension; hypertension; angina pectoris;
 KW myocardial infarction; stroke; ulcer; migraine; vomiting;
 KW psychotic disorder; neurological disorder; anxiety; schizophrenia;
 KW manic depression; bipolar disorder; depression; delirium; dementia;
 KW severe mental retardation; dyskinesia; Parkinson's disease;
 KW Huntington's disease; Gilles de la Tourette's syndrome; gene therapy;
 KW vaccine; drug screening; signal transduction; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200078809-A1.
 XX
 PD 28-DEC-2000.
 XX
 PE 19-JUN-2000; 2000WO-US16869.
 XX
 PR 21-JUN-1999; 99US-0337105.
 XX
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 XX
 PI Elshourbagy N, Shabon U;
 XX
 DR WPI: 2001-102699/11.
 DR N-PSDB: AAC87690.
 XX
 PT New AXOR16, a G-protein coupled receptor polypeptide for screening
 PT agonists and antagonists and for diagnosing and treating microbial
 PT infections, cancer, neurological disorders and asthma
 XX
 NS Claim 1: Page 29-30; 31pp; English.
 XX
 CC The invention relates to the human G protein-coupled receptor AXOR16
 CC (AAB48963) to DNA encoding AXOR16 (AAC87690), and to AXOR16 fragments
 CC and variants. Like all G protein-coupled receptors, AXOR16 has 7
 CC putative transmembrane domains and is involved in signal transduction.
 CC AXOR16 has homology with Gads morhua neuropeptide (NPYm) F and
 CC homology/ structural similarity with Dantio retio neuropeptide Y (NPYRYA).

CC The human AXOR16 gene is located on chromosome 14q12.2. The invention
CC also relates to expression vectors and host cells comprising AXOR16 DNA,
CC to recombinant expression of AXOR16, and to an AXOR16-specific antibody.
CC AXOR16 proteins and nucleotides may be used to treat a wide variety of
CC disorders including bacterial, fungal, protozoal and viral infections,
CC particularly HIV-1 or HIV-2 infections; pain; cancers; benign prostatic
CC hypertrophy; diabetes; obesity; anorexia; bulimia; osteoporosis; asthma;
CC allergies; urinary retention; acute heart failure; hypotension;
CC hypertension; angina pectoris; myocardial infarction; stroke; ulcers;
CC migraine; vomiting; psychotic and neurological disorders such as
CC anxiety, schizophrenia, manic depression, depression, delirium, dementia
CC and severe mental retardation, and dyskinesias, such as Parkinson's
CC disease; Huntington's disease or Gilles de la Tourette's syndrome.
CC AXOR16 proteins, and nucleotides are useful as vaccines, and AXOR16
CC proteins, nucleotides and antibodies may be used in screening compounds
CC for their ability to modulate AXOR16 activity or expression. AXOR16
CC proteins are also useful for inducing an immunological response in a
CC mammal against the above diseases, for antibody production, and to
CC identify membrane bound or soluble receptors for AXOR16. AXOR16
CC nucleotides are also useful as diagnostic reagents and in chromosome
CC localisation and tissue expression studies. The present sequence
CC represents human AXOR16.

SQ Sequence 431 AA;

Query Match	81.88;	Score 27;	DB 22;	Length 431;
Post (read) Similarity	66.78;	Read No. 7	5.03	

Best Local Similarity 66.7%; Pred. NO. 7.5e+02;

Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKMXWR 6

三三三

Search completed: February 27, 2002, 11:41:21
Job time: 460 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:45:25 ; Search time 145.23 Seconds
(without alignments)
3.147 Million cell updates/sec

Title: US-09-446-109A-20

Perfect score: 33

Sequence: 1 FKMxMR 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database : PIR68:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	93.9	75	2	159035
2	31	93.9	989	2	156333
3	29	87.9	81	2	A34019
4	28	84.8	139	2	G71033
5	28	84.8	189	2	B85621
6	28	84.8	245	1	G64834
7	28	84.8	289	2	A84058
8	28	84.8	329	2	A69776
9	28	84.8	459	2	S70702
10	28	84.8	466	2	G69871
11	28	84.8	1087	2	F72283
12	28	84.8	1141	2	A44093
13	28	84.8	1412	2	T01610
14	27	81.8	76	2	A81805
15	27	81.8	94	2	A81071
16	27	81.8	125	2	T02519
17	27	81.8	280	2	T21876
18	27	81.8	338	2	E72264
19	27	81.8	360	2	F81777
20	27	81.8	360	2	E81200
21	27	81.8	417	2	T20327
22	27	81.8	434	2	A96609
23	27	81.8	472	2	G65112
24	27	81.8	472	2	G85985
25	27	81.8	555	2	F84648
26	27	81.8	566	1	B41648
27	27	81.8	603	1	OKAY8W
28	27	81.8	646	2	T36022
29	27	81.8	1099	2	T18257

ALIGNMENTS

30	27	81.8	1397	2	T10466	DNA topoisomerase
31	26	78.8	82	2	T51131	MHC class II beta
32	26	78.8	108	2	F69393	hypothetical prote
33	26	78.8	174	2	S43889	AWL3 protein - wh
34	26	78.8	185	2	G83223	conserved hypothet
35	26	78.8	207	2	E75011	hypothetical prote
36	26	78.8	230	2	I50097	MHC class II beta
37	26	78.8	252	2	H82574	phage-related prot
38	26	78.8	283	2	C83105	formyltetrahydrofo
39	26	78.8	336	2	C69811	nitric-oxide synth
40	26	78.8	345	2	T16935	hypothetical prote
41	26	78.8	347	2	T45551	NADH dehydrogenase
42	26	78.8	362	2	B64527	M-protein - Helico
43	26	78.8	364	2	A64987	hypothetical 40.4
44	26	78.8	364	2	H85856	probable transport
45	26	78.8	374	2	G45252	site-specific reco

RESULT 1
159035
apolipoprotein B p1 - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 02-Jul-1996 #sequence #revision 02-Jul-1996 #text_change 13-Aug-1999
C:Accession: 159035
K:lusis, A.J.; West, R.; Mehraian, M.; Reuben, M.A.; LeBoeuf, R.C.; Kaptein, J.S.; J
Proc. Natl. Acad. Sci. U.S.A. 82, 4597-4601, 1985
A:Title: Cloning and expression of apolipoprotein B, the major protein of low and ver
A:Reference number: 159035; MUID:85270775
A:Accession: 159035
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-75 <RES>
A:Cross-references: GB:M1227; NID:g202954; PIDN:AAA40752.1; PID:g554416
C:Superfamily: apolipoprotein B

Query Match 93.9%; Score 31; DB 2; Length 75;
Best Local Similarity 83.3%; Pred. No. 8.4;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FKMxMR 6
Db 3 FKMxMR 8
RESULT 2
156333
apolipoprotein B - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 26-Jul-1996 #sequence #revision 26-Jul-1996 #text_change 13-Aug-1999
C:Accession: 156333
K:Reuben, M.A.; Svenson, K.L.; Doollittle, M.H.; Johnson, D.F.; Lusis, A.J.; Elvovson,
J. Lipid Res. 29, 1337-1347, 1988
A:Title: Biosynthetic relationships between three rat apolipoprotein B peptides.
A:Reference number: 156333; MUID:89176719
A:Accession: 156333
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-989 <RES>
A:Cross-references: GB:M27440; NID:g623548; PIDN:AAA74690.1; PID:g623549
C:Gene: apob
C:Superfamily: apolipoprotein B

Query Match 93.9%; Score 31; DB 2; Length 989;
Best Local Similarity 83.3%; Pred. No. 88;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 488 FKMWR 493

RESULT 3
A:Accession: A34019

Short neurotoxin B precursor - olive sea snake
C:Species: *Alipysurus laevis* (olive sea snake)
C:Date: 08-Jun-1990 #sequence_revision 08-Jun-1990 #text_change 17-Mar-2000
C:Accession: A34019; S03391
R:Duncan, E.; Guignery-Frelat, G.; Boulain, J.C.; Menez, A.
Toxicol 28, 119-123, 1990
A:Title: Nucleotide sequence and structure analysis of cDNAs encoding short-chain neurotoxin
A:Reference number: A34019; MUID:90232535
A:Accession: A34019

A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-81 <DUC>
A:Cross-references: EMBL:X13373; NID:g62405; PIDN:CAA31748.1; PID:g62406
C:Superfamily: snake toxin
C:Keywords: neurotoxin
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-81/Product: short neurotoxin B #status predicted <MAT>

Query Match 87.9%; Score 29; DB 2; Length 81;
Best Local Similarity 66.7%; Pred. No. 23;
Matches 4; Conservative 1; Mismatches 0; Gaps 0;
OY 1 FKMWR 6
DB 44 YKMTWR 49

RESULT 4
C71033
hypothetical protein PH1559 - *Pyrococcus horikoshii*
C:Species: *Pyrococcus horikoshii*
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C:Accession: C71033
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Seki, M.; Ohnuki, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi, DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaeon
A:Reference number: A71000; MUID:98344137
A:Accession: C71033
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-139 <KAW>
A:Cross-references: GB:AE000006; NID:g3236133; PIDN:BAA30671.1; PID:g3257988
A:Experimental source: strain OT3
A:Note: This accession replaces an interim accession for a sequence replaced by GenBank C:Genetics:
A:Gene: PH1559

Query Match 84.8%; Score 28; DB 2; Length 139;
Best Local Similarity 66.7%; Pred. No. 61;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 1 FKMWR 6
DB 68 FKLNR 73

RESULT 5
B85621
Probable chaperone ycbF [Imported] - *Escherichia coli* (strain O157:H7)
C:Species: *Escherichia coli*
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: B85621
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206531
A:Accession: B85621
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-189 <STO>
A:Cross-references: GB:AE005174; NID:g12514123; PIDN:AGS5430.1; GSPDB:GN00145; UWCP:
A:Experimental source: strain O157:H7, Substrain EDL933
C:Genetics:
A:Gene: ycbF

Query Match 84.8%; Score 28; DB 2; Length 189;
Best Local Similarity 66.7%; Pred. No. 80;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 1 FKMWR 6
DB 149 FKLNR 154

RESULT 6
G64834
Probable fimbrial chaperone precursor - *Escherichia coli*
C:Species: *Escherichia coli*
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: G64834
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of *Escherichia coli* K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: G64834
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-245 <BLAT>
A:Cross-references: GB:AE000196; GB:U00096; NID:g1787169; PIDN:AAC74030.1; PID:g17871
A:Experimental source: strain K-12, Substrain MG1655
C:Genetics:
A:Gene: ycbF
C:Superfamily: chaperone protein papp
C:Keywords: fimbria; membrane protein; molecular chaperone; periplasmic space
F:1-35/Domain: signal sequence #status predicted <SIG>
F:36-245/Product: probable fimbrial chaperone #status predicted <MAT>
F:82-98/Domain: transmembrane #status predicted <TMW>

Query Match 84.8%; Score 28; DB 1; Length 245;
Best Local Similarity 66.7%; Pred. No. 1e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 1 FKMWR 7
DB 149 FKLNR 77

RESULT 7
A84058
Formyltetrahydrofolate deformylase BH3265 [Imported] - *Bacillus halodurans* (strain C-
C:Species: *Bacillus halodurans*
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C:Accession: A84058
R:Takami, H.; Nakasone, K.; Takai, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* a
A:Reference number: A83650; MUID:20263314
A:Accession: A84058
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-289 <STO>
A:Cross-references: GB:AE001518; GB:BA000004; NID:g10175792; PIDN:BA06984.1; GSPDB:G
A:Experimental source: strain C-125

C:Genetics:
A:Gene: BH3265
C:Superfamily: phosphoribosylglycinamide formyltransferase: phosphoribosylglycinamide fo

Query Match 84.8% Score 28; DB 2; Length 289;
Best Local Similarity 66.7% Pred. No. 1.2e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKMxMR 6
1111
DB 80 YKMxMR 85

RESULT 8
A:69776
hypothetical protein yddH [imported] - Bacillus subtilis
C:Species: Bacillus subtilis
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 02-Sep-2000
C:Accession: A69776
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.: Bron, S.; Brouillette, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Erlington, J.; Fabret, C.; Ferrati, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galliz, A.; Gall
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.
Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauel
Y, M.; Ogawa, K.; Ogilawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon
A:Authors: Schleich, S.; Schaefer, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Ser
T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033
A:Accession: A69776
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-329 <KUN>
A:Cross-references: GB:299106; GB:AL009126; NID:92632653; PIDN:CAB12304.1; PID:ell82463;
A:Experimental source: strain 168
C:Genetics:
A:Gene: yddH

Query Match 84.8% Score 28; DB 2; Length 329;
Best Local Similarity 66.7% Pred. No. 1.3e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKMxMR 6
1111
DB 172 FKMxMR 177

RESULT 9
S70702
maacklaln detoxification protein 1 - fungus (Nectria haematococca)
C:Species: Nectria haematococca
C>Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 17-Mar-2000
C:Accession: S70702
R:Coverly, S.F.; Enkerli, J.; Miao, V.P.W.; VanEtten, H.D.
Mol. Gen. Genet. 251, 397-406, 1996
A:Title: A gene for maacklaln detoxification from a dispensable chromosome of Nectria ha
A:Reference number: S70702; MUID:96353814
A:Accession: S70702
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-459 <COV>
A:Cross-references: EMBL:U35892; NID:91019888; PID:91019889
C:Genetics:
A:introns: 60/1; 243/3; 356/2
C:Superfamily: tetracycline 6-hydroxylase

Query Match 84.8% Score 28; DB 2; Length 459;
Best Local Similarity 66.7% Pred. No. 1.8e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKMxMR 6
1111
DB 287 FKLxMR 292

RESULT 10
G69871
hypothetical protein ylaA - Bacillus subtilis
C:Species: Bacillus subtilis
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C:Accession: G69871
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.: Bron, S.; Brouillette, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Erlington, J.; Fabret, C.; Ferrati, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galliz, A.; Gall
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.
Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y, M.; Ogawa, K.; Ogilawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon
A:Authors: Schleich, S.; Schaefer, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
A:Reference number: A69580; MUID:98044033
A:Accession: G69871
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-646 <KUN>
A:Cross-references: GB:299111; GB:AL009126; NID:92633699; PIDN:CAB13344.1; PID:g26338
A:Experimental source: strain 168
C:Genetics:
A:Gene: ylaA
C:Superfamily: Bacillus subtilis hypothetical protein ylaA

Query Match 84.8% Score 28; DB 2; Length 646;
Best Local Similarity 66.7% Pred. No. 2.5e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKMxMR 6
1111
DB 179 FKLxMR 184

RESULT 11
F72283
beta-galactosidase - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: F72283
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Galim, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Cartell, M.M.; Stewart, A.M.; Colton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A:Reference number: A72200; MUID:99287316
A:Accession: F72283
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1087 <ARN>
A:Cross-references: GB:AE001776; GB:AE000512; NID:94981740; PIDN:AD036268.1; PID:g498
C:Genetics:
A:Experimental source: strain MSB8
A:Gene: TMI193

C:Superfamily: beta-galactosidase

Query Match 84.8% Score 28: DB 2: Length 1087;
Best Local Similarity 66.7% Pred. No. 4e+02;
Matches 4: Conservative 1: Mismatches 1: Indels 0: Gaps 0:

OY 1 FKXMR 6
DB 784 FKXMR 789

RESULT 12

A:44093

CGMP-inhibited cAMP phosphodiesterase (EC 3.1.4.-), myocardial form - human

C:Species: Homo sapiens (man)

C:Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 19-May-2000

C:Accession: A44093; S43517

R:Meaccl, E.; Taira, M.; Moos Jr., M.; Smith, C.J.; Movsesian, M.A.; Degerman, E.; Beltr

Proc. Natl. Acad. Sci. U.S.A. 89, 3721-3725, 1992

A:Title: Molecular cloning and expression of human myocardial cGMP-inhibited cAMP phosph

A:Reference number: A44093; MUID:92237240

A:Accession: A44093

A:Molecule type: mRNA

A:Residues: 1-1141 <MEA>

A:Cross-references: GB:M91667; NID:g183392; PIDN:AAA35912.1; PID:g183393

R:Degerman, E.; Moos Jr., M.; Rascon, A.; Vasta, V.; Meaccl, E.; Smith, C.J.; Lindgren,

Biochim. Biophys. Acta 1205, 189-198, 1994

A:Title: Single-step affinity purification, partial structure and properties of human p1

A:Reference number: S43517; MUID:94206994

A:Accession: S43517

A:Status: preliminary

A:Molecule type: protein

A:Residues: 'D', 492, 'X', 494-495, 'P', 497-498, 'YG', 501:661-700:718-759:792-797, 'N', 799-806

A:Genetics: GDB:PDE3A; CGI-PDE

A:Cross-references: GDB:310621

A:Map position: 1p15.2-1p15.1

C:Superfamily: cyclic-nucleotide phosphodiesterase, cGMP-inhibited; 3',5'-cyclic-nucleot

C:Keywords: cGMP binding; phosphoric diester hydrolase

F:751-1019/Domain: 3',5'-cyclic-nucleotide phosphodiesterase homology <CNP>

Query Match 84.8% Score 28: DB 2: Length 1141;
Best Local Similarity 66.7% Pred. No. 4.2e+02;
Matches 4: Conservative 1: Mismatches 1: Indels 0: Gaps 0:

OY 1 FKXMR 6
DB 232 FKXMR 237

RESULT 13

A:44093

RNA-directed DNA polymerase homolog At2g18820 - Arabidopsis thaliana

N:Alternate names: RNA-directed DNA polymerase homolog F19F24.2

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 02-Feb-2001

C:Accession: T01610; A84569

R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul

submitted to the EMBL Data Library, April 1998

A:Description: Arabidopsis thaliana chromosome II BAC F19F24 genomic sequence.

A:Reference number: Z14153

A:Accession: T01610

A:Status: translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-1412 <ROU>

A:Cross-references: EMBL:AC003673; NID:g3004543; PID:g3004545

A:Experimental source: cultivar Columbia

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanden, S.E.; Unayam, L.; Tallon, L.

cuss, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487

A:Accession: A84569

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1412 <SNO>

A:Cross-references: GB:AE002093; NID:g4185148; PIDN:AD08951.1; GSPDB:GN00139

C:Genetics: A:Gene: At2g18820; F19F24.2

A:Map position: 2

A:Introns: 175/1; 253/1; 332/3; 647/3; 910/1; 1169/1

Query Match 84.8% Score 28: DB 2: Length 1412;
Best Local Similarity 66.7% Pred. No. 5e+02;
Matches 4: Conservative 1: Mismatches 1: Indels 0: Gaps 0:

OY 1 FKXMR 6
DB 1143 FKXMR 1148

RESULT 14

A:81805

Hypothetical protein NMA1795 [imported] - Neisseria meningitidis (strain 22491 serogr

C:Species: Neisseria meningitidis

C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001

C:Accession: A81805

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo

Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491

A:Reference number: A81775; MUID:20222556

A:Accession: A81805

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-776 <PAR>

A:Cross-references: GB:AL162757; GB:AL157959; NID:g7380371; PIDN:CA85021.1; PID:g738

A:Experimental source: serogroup A, strain 22491

C:Genetics: A:Gene: NMA1795

Query Match 81.8% Score 27: DB 2: Length 76;
Best Local Similarity 80.0% Pred. No. 56;
Matches 4: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

OY 1 FKXMR 5
DB 65 FKXMR 69

RESULT 15

A:81805

Hypothetical protein NMA1546, NMA1631 [imported] - Neisseria meningitidis (strain MC5

C:Species: Neisseria meningitidis

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001

C:Accession: E81071; A81061

R:Petzelt, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,

Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.

Science 287, 1809-1815, 2000

A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappelli, R.;

A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

A:Reference number: A81000; MUID:20175755

A:Accession: E81071

A:Molecule type: DNA

A:Residues: 1-94 <TET>

A:Cross-references: GB:AE002504; GB:AE002098; NID:g7226785; PIDN:AAE41901.1; PID:g722

A:Experimental source: serogroup B, strain MC58

A:Accession: A81061

A:Molecule type: DNA

A:Residues: 1-94 <TE2>
A:Cross-references: GB:AE002513; GB:AE002098; NID:97226875; PIDN:AAF1981.1; PID:9722688
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1546; NMB1631

Query Match	81.8%	Score 27	DB 2	Length 94
Best Local Similarity	80.0%	Pred. No. 68		
Matches	4	Conservative	0	Mismatches 1; Indels 0; Gaps 0
QY	1	FKMXW	5	
	111-1			
DB	83	FKMAW	87	

Search completed: February 27, 2002, 11:45:26
Job time: 705 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:42:50 : Search time 78.39 Seconds
(without alignments)
2.806 Million cell updates/sec

Title: US-09-446-109a-20

Perfect score: 33
Sequence: 1 FKXMR 6

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	87.9	81	1 NXS2_AIPLA	P19959 alipysurus 1
2	28	84.8	236	1 YCBF_ECOLI	P40876 escherichia
3	28	84.8	1084	1 BGAL_THEMA	O56307 thermotoga
4	28	84.8	1141	1 CN3A_HUMAN	Q14432 homo sapien
5	28	84.8	1141	1 CN3A_RAT	Q62865 ratus norv
6	27	81.8	338	1 YD49_THEMA	O9X170 thermotoga
7	27	81.8	471	1 GLTD_ECOLI	P09832 escherichia
8	27	81.8	566	1 MP12_HUMAN	P30305 homo sapien
9	27	81.8	603	1 SAT4_YEAST	P25333 saccharomyc
10	27	81.8	646	1 YE14_SCHPO	O13659 schizosacch
11	27	81.8	1099	1 PLCL_CANAL	O13433 candida alb
12	27	81.8	1398	1 TOP2_PLAFK	P41001 plasmodium
13	26	78.8	108	1 YB51_ARCFU	O29114 archaeoglob
14	26	78.8	294	1 PT54_SACBA	O13364 saccharomyc
15	26	78.8	364	1 YE3B_ECOLI	P33314 escherichia
16	26	78.8	499	1 PROD_ARATH	P92983 arabidopsis
17	26	78.8	618	1 SPPA_ECOLI	P08395 escherichia
18	26	78.8	831	1 TVID_SALTI	O04974 salmonella
19	26	78.8	895	1 POL_RSVF	P03354 rous sarcom
20	26	78.8	988	1 CHS1_EXODE	P30600 exophiala d
21	26	78.8	1021	1 MANA_RHOMR	P49425 rhodothermu
22	26	78.8	1138	1 DPS2_YEAST	P53037 saccharomyc
23	26	78.8	1144	1 NOS2_MOUSE	P29477 mus musculu
24	26	78.8	1147	1 NOS2_RAT	O66518 ratus norv
25	26	78.8	1147	1 NS2D_HUMAN	O60591 homo sapien
26	26	78.8	1246	1 VP03_HSVSA	O01000 herpesviru
27	26	78.8	1564	1 PDRA_YEAST	P51533 saccharomyc
28	25	75.8	66	1 CAD0_BUNMU	P81783 bungarus ca
29	25	75.8	87	1 NXLH_BUNMU	P15818 bungarus mu
30	25	75.8	198	1 SARI_TOHAC	P52885 nicotiana t
31	25	75.8	212	1 COAT_PEBV	P14849 pea early b
32	25	75.8	251	1 Y29K_SSVI	P20202 sulfolobus
33	25	75.8	262	1 LAM5_HUMAN	O13571 homo sapien

34	25	75.8	287	1 SUD2_ARCFU	O28098 archaeoglob
35	25	75.8	299	1 YTAI_BACSU	P37513 bacillus su
36	25	75.8	347	1 NU2M_FELCA	P48905 felis silve
37	25	75.8	384	1 RC11_ECOLI	P10487 escherichia
38	25	75.8	384	1 RC12_ECOLI	P16470 escherichia
39	25	75.8	396	1 TL3A_MOUSE	P26952 mus musculu
40	25	75.8	417	1 YNFM_ECOLI	P43531 escherichia
41	25	75.8	428	1 YE63_SCHPO	O14249 schizosacch
42	25	75.8	432	1 REBX_SAITI	O99191 salmonella
43	25	75.8	433	1 YMN3_PODAN	P15563 podopora a
44	25	75.8	438	1 SYD_METUA	O58950 methanococc
45	25	75.8	483	1 BIAR_MELGA	P07700 melaleucis 9

ALIGNMENTS

```

RESULT 1
ID NX_S2_AIPLA STANDARD: PRT: 81 AA.
AC P19959: P01439:
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 22, Last sequence update)
DT 01-MAY-1993 (Rel. 22, Last annotation update)
DE SHORT NEUROTOXIN B PRECURSOR.
GS Alipysurus laevis (Olive sea snake).
CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC Lepidosauria; Squamata; ScleroGLOSSa; Serpentes; Colubroidea;
CC Elapidae; Hydrophiinae; Alipysurus.
OX NCBI_TaxID=8678:
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Venom gland;
EA MEDLINE=90232535; PubMed=2330602;
FA Duncanel F., Guignery-Frelat G., Boulain J.-C., Menez A.;
RT "Nucleotide sequence and structure analysis of cDNAs encoding short-
chain neurotoxins from venom glands of a sea snake (Alipysurus
laevis).";
RL Toxicon 28:119-123(1990).
NN (2)
HP SEQUENCE OF 22-81.
KC TISSUE=Venom;
FX MEDLINE=76160626; PubMed=1259717;
FA Maeda N., Tamiya N.;
RT "Isolation, properties and amino acid sequences of three neurotoxins
from the venom of a sea snake, Alipysurus laevis.";
RL Biochem. J. 153:79-87(1976).
CC -!- FUNCTION: PRODUCES PERIPHERAL PARALYSIS BY BLOCKING NEUROMUSCULAR
TRANSMISSION AT THE POSTSYNAPTIC SITE. BINDS TO THE NICOTINIC
ACETYLCHOLINE RECEPTOR.
CC -!- MISCELLANEOUS: LD(50) IS 0.067 MG/KG BY INTRAMUSCULAR INJECTION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X13373; CAA1748.1; -
DR PIR: A01707; MIAV1.
DR PIR: S03391; S03391.
DR PIR: A34019; A34019.
DR HSSP: P01426; INFA.
DR InterPro: IPR003571; Snake_toxin.
DR Pfam: PF00067; Toxin_1.
DR ProDom: PD000206; Snake_toxin_1.
DR PROSITE: PS00272; SNAKE_TOXIN_1.
KW Venom; Neurotoxin; Postsynaptic neurotoxin; Multigene family; Signal.
FT SIGNAL 1 21
FT CHAIN 22 81 SHORT NEUROTOXIN B.
FT PT 24 43 BY SIMILARITY.
FT DISULFID 24

```

```
FT DISULFID 38 60 BY SIMILARITY.
CC DISULFID 62 73 BY SIMILARITY.
FT DISULFID 74 79 BY SIMILARITY.
SQ SEQUENCE 81 AA: 9047 MW: E9D6403A5D59CE43 CRC64:

Query Match      87.9% Score 29: DB 1: Length 81:
Best Local Similarity 66.7% Pred. No. 11:
Matches 4: Conservative 1: Mismatches 1: Indels 0: Gaps 0:

OY 1 FKMXMR 6
DB 44 YKMTWR 49

RESULT 2
ID YCBF_ECOLI STANDARD: PRT: 236 AA.
AC P40876; P75861;
DT 01-FEB-1995 (Rel. 31, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL FIBRILLAL CHAPERONE YCBF PRECURSOR.
CN YCBF OR B0944.
OS Escherichia coli.
OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick II.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RA Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiiuchi T.;
RA "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RA DNA Res. 3:137-155(1996).
RN [3]
RP SEQUENCE OF 165-236 FROM N.A.
RX MEDLINE=85285014; PubMed=2992959;
RA Larsen N.J., Jensen K.F.;
RT "Nucleotide sequence of the pyrd gene of Escherichia coli and
RT characterization of the flavoprotein dihydrofolate dehydrogenase.";
RA Eur. J. Biochem. 151:59-65(1985).
RN [4]
RP IDENTIFICATION.
RX MEDLINE=95075659; PubMed=7984428;
RA Borodovsky M., Rudd K.E., Koonin E.V.;
RT "Intrinsic and extrinsic approaches for detecting genes in a
RT bacterial genome.";
RA Nucleic Acids Res. 22:4756-4767(1994).
CC -1- SUBCELLULAR LOCATION: PERIPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PERIPLASMIC PILUS CHAPERONE FAMILY.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000196; AAC74030.1; ALT_INIT.
DR EMBL: D90732; BAA35699.1; ALT_INIT.
DR EMBL: X02826; -; NOT_ANNOTATED_CDS.
DR EcGene: EG12397; ycbf.
DR InterPro: IPR001829; PIL1_Chaperone.
DR Pfam: PF00345; PIL1_assembly; 1.
DR PRINTS: PR00969; CHAPERONPIL1.
DR ProDom: PD001447; PIL1_chaperone; 1.
DR ProSite: PS00635; PIL1_CHAPERONE; FALSE_NEG.
KW Hypothetical protein: Chaperone; Fimbria; Periplasmic; Signal;
KW Immunoglobulin domain; Complete proteome.
FT SIGNAL 1 26
FT CHAIN 27 236
FT CHAIN 27 236 HYPOTHETICAL FIBRILLAL CHAPERONE YCBF.
SQ SEQUENCE 236 AA: 25853 MW: 19B8101EDCFE5607 CRC64:

Query Match      84.8% Score 28: DB 1: Length 236:
Best Local Similarity 66.7% Pred. No. 48:
Matches 4: Conservative 1: Mismatches 1: Indels 0: Gaps 0:

OY 1 FKMXMR 6
DB 140 FKLFWR 145

RESULT 3
ID BGAL_THEMA STANDARD: PRT: 1084 AA.
AC Q56307; Q33834;
DT 15-DEC-1998 (Rel. 37, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE BETA-GALACTOSIDASE (EC 3.2.1.23) (LACTASE).
CN LACZ OR TM1193.
OS Thermotoga maritima.
OC Bacteria: Thermotogales: Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MSB8 / DSM 3109;
RX MEDLINE=94374690; PubMed=8088532;
RA Moore J.B., Markiewicz P., Miller J.H.;
RT "Identification and sequencing of the Thermotoga maritima lacz gene,
RT part of a divergently transcribed operon.";
RT Gene 147:101-106(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RT Nature 399:323-329(1999).
RN [3]
RP SEQUENCE OF 554-1084 FROM N.A.
RX STRAIN-MSB8 / DSM 3109;
RX MEDLINE=98413539; PubMed=9741105;
RA Liebl W., Wagner B., Schellhase J.;
RT "Properties of an alpha-galactosidase, and structure of its gene
RT galA, within an alpha and beta-galactoside utilization gene cluster
RT of the hyperthermophilic bacterium Thermotoga maritima.";
RT Syst. Appl. Microbiol. 21:1-11(1998).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
CC GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
```

```

CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U08186; AAA50597.1; -.
CC EMBL: AF001776; AAD36268.1; ALT_INIT.
CC EMBL: A001072; CAA04513.1; -.
CC TIGR: TM1193; -.
CC DR InterPro: IPR001649; Glyco_hydro_2.
CC DR Pfam: PF00703; Glyco_hydro_2; 1.
CC DR PRINTS: PR00132; GLYHYDRLASE2.
CC DR PROSITE: PS00719; GLYCOSYL_HYDROL_F2_1; 1.
CC DR PROSITE: PS00608; GLYCOSYL_HYDROL_F2_2; 1.
CC KW Hydrolase; Glycosidase; Complete proteome.
CC FT ACT_SITE 441 441 PROTON DONOR (BY SIMILARITY).
CC FT ACT_SITE 507 507 NUCLEOPHILE (BY SIMILARITY).
CC FT CONFLICT 152 175 SKDSCPAEFLTDVLRPKNLIT -> RQQLHARRIQTH
CC FT CONFLICT 152 175 RSKRTRESDH (IN REF. 1).
CC FT CONFLICT 1028 1084 SEKQVVLFDGNEYSRRVVRPPKKEELVFKVEGLKGE
CC FT CONFLICT 1028 1084 HLHHTNLNTRKTIYVR -> RQCKTGSLC (IN REF.
CC FT SEQUENCE 1084 AA; 127607 MW; D52E3B762B53DFFC CRC64;
CC SQ

```

```

RC TISSUE-Penis;
RA Kuthe A., Eckel H., Magert H.J., Uckert S., Forssmann W.G., Jonas U.,
RA Stief C.G.;
RT "Molecular characterization of phosphodiesterases in human Corpus
RT cavernosum.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: GUANOSINE 3',5'-CYCLIC PHOSPHATE + H(2)O =
CC GUANOSINE 5'-PHOSPHATE.
CC -1- ENZYME REGULATION: INHIBITED BY cGMP.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE PHOSPHODIESTERASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M91667; AAA35912.1; -.
CC EMBL: U36798; AAB18673.1; -.
CC EMBL: AJ005036; CAA06304.1; -.
CC MIM: 123805; -.
CC DR InterPro: IPR003607; HDC.
CC DR InterPro: IPR002073; PDEase.
CC DR Pfam: PF00233; PDEase; 1.
CC DR SMART: SM00471; HDC; 1.
CC DR PROSITE: PS00126; PDEASE_L; 1.
CC KW Hydrolase; cGMP; Membrane.
CC FT DOMAIN 95 98 POLY-ALA.
CC FT DOMAIN 99 102 POLY-GLU.
CC FT DOMAIN 288 291 POLY-ARG.
CC FT DOMAIN 440 445 POLY-THR.
CC FT DOMAIN 870 873 POLY-ALA.
CC FT DOMAIN 1040 1045 POLY-GLU.
CC FT DOMAIN 1121 1125 POLY-GLU.
CC FT CONFLICT 12 12 N -> D (IN REF. 3).
CC FT CONFLICT 63 64 SA -> LP (IN REF. 1).
CC FT CONFLICT 69 69 S -> C (IN REF. 2).
CC FT CONFLICT 110 110 G -> A (IN REF. 2).
CC FT SEQUENCE 1141 AA; 125107 MW; F480B9B981368484 CRC64;
CC SQ

```

```

Query Match          84.8%; Score 28; DB 1; Length 1084;
Best Local Similarity 66.7%; Pred. No. 2e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKXWR 6
1111
DB 781 FKXWR 786

RESULT 4
CN3A_HUMAN STANDARD; PRT; 1141 AA.
AC Q14432; Q13348; 060865;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE cGMP-INHIBITED 3',5'-CYCLIC PHOSPHODIESTERASE A (EC 3.1.4.17) (CYCLIC
DE cMP INHIBITED PHOSPHODIESTERASE A) (CGI-PDE A).
DE PDE3A.
CN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE-Heart;
RX MEDLINE=92237240; PubMed=1315035;
RA Meacci E., Taira M., Moos M. Jr., Smith C.J., Movsesian M.A.,
RA Degerman E., Belfrage P., Manganiello V.;
RT "Molecular cloning and expression of human myocardial cGMP-inhibited
RT cAMP phosphodiesterase.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:3721-3725(1992).
RN 12
RP SEQUENCE FROM N.A.
RC TISSUE-Blood;
RX MEDLINE=96329565; PubMed=8695850;
RA Cheung P.P., Xu H., McLaughlin M.W., Ghazaleh F.A., Livi G.P.,
RA Colman R.W.;
RT "Human platelet cG-PDE: expression in yeast and localization of the
RT catalytic domain by deletion mutagenesis.";
RL Blood 88:1321-1329(1996).
RN 13
RP SEQUENCE FROM N.A.

```

```

Query Match          84.8%; Score 28; DB 1; Length 1141;
Best Local Similarity 66.7%; Pred. No. 2.1e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKXWR 6
1111
DB 232 FKXWR 237

RESULT 5
CN3A_RAT STANDARD; PRT; 1141 AA.
AC Q62865;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE cGMP-INHIBITED 3',5'-CYCLIC PHOSPHODIESTERASE A (EC 3.1.4.17) (CYCLIC
DE cMP INHIBITED PHOSPHODIESTERASE A) (CGI-PDE A).
DE PDE3A.
CN Rattus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE-Adipose tissue;
RX MEDLINE=98294701; PubMed=9631240;
RA He R., Komar N., Ekholm D., Murata T., Taira M., Hockman S.C.,

```

RA Degerman E., Manganiello V.C.;
 RT *Expression and characterization of deletion recombinants of two
 RL CGMP-Inhibited cyclic nucleotide phosphodiesterases (PDE-3)."
 CC Cell Biochem. Biophys. 29:89-111(1998).
 CC -1- CATALYTIC ACTIVITY: GUANOSINE 3',5'-CYCLIC PHOSPHATE + H(2)O =
 CC GUANOSINE 5'-PHOSPHATE.
 CC -1- ENZYME REGULATION: INHIBITED BY CGMP.
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE PHOSPHODIESTERASE
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: U38179; AAA84964.1; -
 CC DR InterPro: IPR003607; HDC.
 CC DR InterPro: IPR02073; PDEase.
 CC DR Pfam: PF00233; PDEase. 1.
 CC DR SMART: SM00471; HDC. 1.
 CC DR PROSITE: PS00126; PDEASE.L; 1.
 CC KM Hydrolase: CGMP; Membrane.
 CC SO SEQUENCE: 1141 AA; 124300 MW; A333DFB446F33F3 CRC64;

Query Match 84.8%; Score 28; DB 1; Length 1141;
 Best Local Similarity 66.7%; Pred. No. 2.1e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FRKXMR 6
 11: 11
 DB 229 FRVAMR 234

RESULT 6
 YD49_THEME STANDARD: PRT; 338 AA.
 ID YD49_THEME
 AC O9X170;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 GN HYPOTHEICAL PROTEIN TMI349.
 GN TMI349.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogales; Thermotoga.
 OX NCBI_TaxId=2336;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=MSB8 / DSM 3109;
 RC MEDLINE=99287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback F.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT *Evidence for lateral gene transfer between Archaea and Bacteria from
 RT genome sequence of Thermotoga maritima."
 RL Nature 399:322-329(1999).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE UPF0118 (PERM) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC EMBL: AE001789; AAD36420.1; -
 DR TIGR: TMI349; -
 DR InterPro: IPR002549; UPF0118.
 CC Pfam: PF01594; UPF0118; 1.
 CC KM Hypothetical protein: Transmembrane; Complete proteome.
 CC -1- TRANSMEM 20 40
 CC FT TRANSMEM 68 88 POTENTIAL.
 CC FT TRANSMEM 147 167 POTENTIAL.
 CC FT TRANSMEM 203 223 POTENTIAL.
 CC FT TRANSMEM 239 259 POTENTIAL.
 CC FT TRANSMEM 263 283 POTENTIAL.
 CC FT TRANSMEM 297 317 POTENTIAL.
 CC SQ SEQUENCE 338 AA; 39386 MW; 9B21A19682078AD CRC64;

Query Match 81.8%; Score 27; DB 1; Length 338;
 Best Local Similarity 66.7%; Pred. No. 1.1e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FRKXMR 6
 11: 11
 DB 323 FRVAMR 328

RESULT 7
 GLTD_ECOLI STANDARD: PRT; 471 AA.
 ID GLTD_ECOLI
 AC P09832;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE GLUTAMATE SYNTHASE (NADPH) SMALL CHAIN (EC 1.4.1.13) (GLUTAMATE
 DE SYNTHASE BETA SUBUNIT) (NADPH-GOGAT) (GLTS BETA CHAIN).
 GN GLTD OR ASPB OR B3213.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxId=562;
 RN (1)
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=K12;
 RC MEDLINE=88152492; PubMed=3326786;
 RA Oliver G., Gossel G., Sanchez-Pescador R., Lozoya E., Ku L.M.,
 RA Flores N., Becerril B., Valle F., Bolivar F.;
 RT *Determination of the nucleotide sequence for the glutamate synthase
 RT structural genes of Escherichia coli K-12.";
 RT Gene 60:1-11(1987).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RC MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT *The complete genome sequence of Escherichia coli K-12.";
 RT Science 277:1453-1474(1997).
 RN (3)
 RP SEQUENCE OF 1-12.
 RC STRAIN=K12 / EMC2;
 RC MEDLINE=97443975; PubMed=9298646;
 RA Link A.J., Robinson K., Church G.M.;
 RT *Comparing the predicted and observed properties of proteins encoded
 RT in the genome of Escherichia coli K-12.";
 RT Electrophoresis 18:1259-1313(1997).
 RN (4)
 RP DISCUSSION OF SEQUENCE.
 RC MEDLINE=89098858; PubMed=2643092;
 RA Gossel G., Merino E., Recillas F., Oliver G., Becerril B., Bolivar F.;
 RT *Amino acid sequence analysis of the glutamate synthase enzyme from
 RT Escherichia coli K-12.";
 RT Protein Seq. Data Anal. 2:9-16(1989).

CC -1- CATALYTIC ACTIVITY: 2 L-GLUTAMATE + NADP(+) = L-GLUTAMINE +
 CC 2-OXOGLUTARATE + NADPH.
 CC -1- COFACTOR: BINDS A 4FE-4S CLUSTER.
 CC -1- PATHWAY: NITROGEN METABOLISM; GLUTAMATE BIOSYNTHESIS.
 CC THE CATALYZED REACTION BRINGS TOGETHER THE NITROGEN AND
 CC CARBON METABOLISM.
 CC -1- SUBUNIT: AGGREGATE OF 4 CATALYTICAL ACTIVE HETERODIMERS,
 CC CONSISTING OF A LARGE AND A SMALL SUBUNIT.
 CC -1- MISCELLANEOUS: GLUTAMINE BINDS TO THE LARGE SUBUNIT AND TRANSFERS
 CC THE AMIDO GROUP TO 2-OXO-GLUTAMATE THAT APPARENTLY BINDS TO THE
 CC SMALL SUBUNIT.
 CC -1- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF
 CC BACTERIAL-TYPE 4FE-4S FERREDOXINS.
 CC -1- SIMILARITY: STRONG, TO E.COLI AEGA AND YGFT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: M18747; AAA23905.1; -
 CC DR EMBL: U18997; AAA58015.1; -
 CC DR EMBL: AE000400; AAC76245.1; -
 CC DR PIR: B29617; B29617.
 CC DR SWISS-2DPAGE: P09832; COLI.
 CC DR EC02DBASE: F050.4; 6TH EDITION.
 CC DR EC02BASE: EC0404.4; g1td.
 CC DR Ecocore: EC0404.4; g1td.
 CC DR InterPro: IPR000759; Adnrx_redctase.
 CC DR InterPro: IPR001327; FAD_pyr_redox.
 CC DR InterPro: IPR000205; NAD_binding.
 CC DR Pfam: PF00070; pyr_redox.1.
 CC DR PRINTS: PR000419; ADXRDTASE.
 CC DR Oxidoreductase: Glutamate biosynthesis: Iron-sulfur; 4Fe-4S; NADP;
 CC Complete proteome.
 CC KM INIT_MEN 0
 CC FT METAL 93 93 IRON-SULFUR (4FE-4S) (POTENTIAL).
 CC FT METAL 97 97 IRON-SULFUR (4FE-4S) (POTENTIAL).
 CC FT METAL 103 103 IRON-SULFUR (4FE-4S) (POTENTIAL).
 CC FT METAL 107 107 IRON-SULFUR (4FE-4S) (POTENTIAL).
 CC FT CONFLICT 37 50 GQAKADRCCLSCG -> ARPKRLTAACRAA (IN
 CC REF. 1).
 CC FT CONFLICT 122 122 E -> K (IN REF. 1).
 CC FT CONFLICT 173 173 V -> C (IN REF. 1).
 CC FT CONFLICT 226 262 VYALPEPLANTKO -> CTGCRSSSPTPNS (IN
 CC REF. 1).
 CC FT CONFLICT 311 312 KH -> ND (IN REF. 1).
 CC FT CONFLICT 375 399 GRRRAEIVAGSEHIVPADAVIMAFG -> ASPRGDRCRPRT
 CC YRTGCGDHGVN (IN REF. 1).
 CC FT SEQUENCE 471 AA: 51884 MW: C4070F509C561A8 CRC64:
 CC
 CC Query Match 81.8%; Score 27; DB 1; Length 471;
 CC Best Local Similarity 66.7%; Pred. No. 1.4e+02;
 CC Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC QY 1 FKKXMR 6
 CC DB 130 FEMGWR 135
 CC
 CC RESULT 8
 CC MPT2_HUMAN STANDARD; PRT; 566 AA.
 CC AC P30305;
 CC DT 01-APR-1993 (Rel. 25, Created)
 CC DT 01-APR-1993 (Rel. 25, Last sequence update)
 CC DT 20-AUG-2001 (Rel. 40, Last annotation update)
 CC DE M-PHASE INDUCER PHOSPHATASE 2 (EC 3.1.3.48) (DUAL SPECIFICITY
 CC DE PHOSPHATASE CDC25B).
 CC GN CDC25B OR CDC25HJ2.

CC US Homo sapiens (Human).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CC OX NCBI_TaxID:9606;
 CC RN [1]
 CC RN SEQUENCE FROM N.A.
 CC RX MEDLINE=92103683; PubMed=1836978;
 CC RA Galaktionov K.I., Beach D.;
 CC RT "Specific activation of cdc25 tyrosine phosphatases by B-type
 CC RT cyclins: evidence for multiple roles of mitotic cyclins.";
 CC RL Cell 67:1181-1194(1991).
 CC RN [2]
 CC RN SEQUENCE FROM N.A.
 CC RX MEDLINE=92118716; PubMed=1662986;
 CC RA Nagata A., Igarashi M., Jinno S., Suto K., Okayama H.;
 CC RT "An additional homolog of the fission yeast cdc25+ gene occurs in
 CC RT humans and is highly expressed in some cancer cells.";
 CC RN New Biol. 3:959-968(1991).
 CC RN [3]
 CC RN SEQUENCE OF 130-191 FROM N.A.
 CC RP Baldwin V., Ducommun B.;
 CC RA Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
 CC RN [4]
 CC RP X-RAY CRYSTALLOGRAPHY (1.91 ANGSTROMS) OF 356-566.
 CC RX MEDLINE=20013068; PubMed=10543950;
 CC RA Reynolds R.A., Yem A.W., Wolfe C.L., Deibel M.R. Jr., Chidester C.G.,
 CC RA Walenpaugh K.D.;
 CC RT "Crystal structure of the catalytic subunit of Cdc25B required for
 CC RT G2/M phase transition of the cell cycle.";
 CC RL J. Mol. Biol. 293:559-568(1999).
 CC CC -1- FUNCTION: THIS PROTEIN FUNCTIONS AS A DOSAGE-DEPENDENT INDUCER IN
 CC CC MITOTIC CONTROL. IT IS A TYROSINE PROTEIN PHOSPHATASE REQUIRED FOR
 CC CC PROGRESSION OF THE CELL CYCLE. IT MAY DIRECTLY DEPHOSPHORYLATE
 CC CC P34(CDC2) AND ACTIVATE THE P34(CDC2) KINASE ACTIVITY.
 CC CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
 CC CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
 CC CC -1- ENZYME REGULATION: STIMULATED BY CYCLINS B
 CC CC AND IN GENERAL TO PROTEIN-TYROSINE PHOSPHATASES.
 CC CC -1- SIMILARITY: STRONG, TO OTHER SPECIES M-PHASE INDUCER PHOSPHATASE
 CC AND IN GENERAL TO PROTEIN-TYROSINE PHOSPHATASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: M81934; AAA58416.1; -
 CC DR EMBL: S78187; AAH21139.1; -
 CC DR EMBL: X96436; CAA65303.1; -
 CC DR PIR: B41648; B41648.
 CC DR PDB: 10B0; 29-APR-00.
 CC DR PDB: 1CWR; 28-AUG-00.
 CC DR MIM: 116949; -
 CC DR InterPro: IPR000751; MPT_Phoiphase.
 CC DR InterPro: IPR001763; Rhodanese_domain.
 CC DR Pfam: PF00581; Rhodanese.1.
 CC DR PRINTS: PR00716; MPTPHPTASE.
 CC DR SMART: SM00450; RHOD.1.
 CC DR Cell division; Mitosis; Hydrolase; Multigene family; 3D-structure.
 CC KW ACT_SITE 473 473
 CC FT CONFLICT 561 561 S -> D (IN REF. 2).
 CC FT SEQUENCE 566 AA: 63442 MW: 7C54E3E799ECD450 CRC64:
 CC SQ
 CC Query Match 81.8%; Score 27; DB 1; Length 566;
 CC Best Local Similarity 66.7%; Pred. No. 1.7e+02;
 CC Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC QY 1 FKKXMR 6
 CC DB 185 FKKPMK 190

```

RESULT 9
SAT4_YEAST STANDARD: PRT: 603 AA.
ID SAT4_YEAST
AC P25333:
DT 01-MAY-1992 (rel. 22, Last sequence update)
DT 01-MAY-1992 (rel. 37, Last annotation update)
DE 15-DEC-1998 (rel. 37, Last annotation update)
SERIAL/THREONINE-PROTEIN KINASE SAT4 (EC 2.7.1.-).
GN SAT4 OR YCR008M OR YCR8M OR YCR101 OR YCR046.
OS Saccharomyces cerevisiae (baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
  Saccharomycetiales; Saccharomycetaceae; Saccharomyces.
NCBI_TaxID=4932;
RX MEDLINE=92116650; PubMed=1767593;
RA Skala J., Purnelle B., Crouzet M., Algle M., Goffeau A.;
RT "The open reading frame YCR101 located on chromosome III from
  Saccharomyces cerevisiae is a putative protein kinase."
  Yeast 7:651-655(1991).
  [2]
RX MEDLINE=92254505; PubMed=1580102;
RA Bileau N., Fremaux C., Hebrard S., Menara A., Algle M., Crouzet M.;
RT "The complete sequence of a 10.8kb fragment to the right of the
  chromosome III centromere of Saccharomyces cerevisiae."
  Yeast 8:61-70(1992).
  [1]
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (see http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
  [1]
CC
DR EMBL: S76380; AAB20894.1; -
DR EMBL: 211114; CAAT7445.1; -
DR EMBL: X59720; CAA42325.1; -
DR PIR: S19505; OKBY8M.
DR PIR: S17470; S17470.
DR HSP: 063450; 1A06.
DR SGD: S0000601; SAT4.
DR InterPro: IPR000719; Euk_pkinase.
DR pfam: PF00069; pkinase.1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP.1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST.1.
DR TRANSFERASE: Serine/threonine-protein kinase; ATP-binding.
KW DOMAIN 99 111 SER-RICH.
FT DOMAIN 316 590 PROTEIN KINASE.
FT NP_BIND 322 330 ATP (BY SIMILARITY).
FT BINDING 353 353 ATP (BY SIMILARITY).
FT ACT_SITE 449 449 BY SIMILARITY.
SQ SEQUENCE 603 AA: 66665 MW: 9DD31B74C05EE212 CRC64;

```

```

AC 013869:
DT 15-JUL-1998 (rel. 36, Created)
DT 15-JUL-1998 (rel. 36, Last sequence update)
DT 15-DEC-1998 (rel. 37, Last annotation update)
DE PUTATIVE GTP-BINDING PROTEIN C1B3.04C.
GN SPAC1B3.04C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
  Schizosaccharomycetiales; Schizosaccharomycetaceae;
  Schizosaccharomyces.
OC NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972.
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A., Wood V.;
  Submitted (AUG-1997) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
  LEPA SUBFAMILY.
  [1]
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (see http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
  [1]
CC
DR EMBL: Z98598; CAB11233.1; -
DR HSP: P13551; IELO.
DR InterPro: IPR000795; GTP_EFTU.
DR pfam: PF00009; GTP_EFTU.1.
DR PROSITE: PS00301; EFATOR_GTP; FALSE_NEG.
KW Hypothetical protein; GTP-binding.
FT NP_BIND 65 72 GTP (POTENTIAL).
FT NP_BIND 130 134 GTP (POTENTIAL).
FT NP_BIND 184 187 GTP (POTENTIAL).
SQ SEQUENCE 646 AA: 72683 MW: F9FA9498D384503E CRC64;

```

Query Match 81.8%; Score 27; DB 1; Length 646;
 Best Local Similarity 66.7%; Pred. No. 1.9e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

OY 1 FKNXR 6
DB 21 FKIRWR 26

```

RESULT 11
 PCL1_CANAL STANDARD: PRT: 1099 AA.
 ID PCL1_CANAL
 AC 013433;
 DT 15-JUL-1999 (rel. 38, Created)
 DT 15-JUL-1999 (rel. 38, Last sequence update)
 DT 15-JUL-1999 (rel. 38, Last annotation update)
 DE 1-PROSPHATIDYLINOSITOL-4,5-BISPHOSPHATE PHOSPHODIESTERASE 1
 (EC 3.1.4.11) (PDC-1) (PHOSPHOLIPASE C-1).
 GN PCL1.
 OS Candida albicans (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetiales; mitosporic Saccharomycetiales; Candida.
 NCBI_TaxID=5476;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=132A;
 RX MEDLINE=98129081; PubMed=9467900;
 RA Bennett D.E., McCreary C.E., Coleman D.C.;
 "Genetic characterization of a phospholipase C gene from Candida
 albicans: presence of homologous sequences in Candida species other
 than Candida albicans."
 Microbiology 144:55-72(1998).
 CC -1- FUNCTION: THE PRODUCTION OF THE SECOND MESSENGER MOLECULES
 DIACYLGLYCEROL (DAG) AND INOSITOL 1,4,5-TRISPHOSPHATE (IP3) IS


```

MEDATED BY ACTIVATED PHOSPHATIDYLINOSITOL-SPECIFIC PHOSPHOLIPASE
C ENZYMES.
-1- CATALYTIC ACTIVITY: 1-PHOSPHATIDYL-D-MYO-INOSITOL 4,5-BISPHOSPHATE
+ H(2)O = D-MYO-INOSITOL 1,4,5-TRIPHOSPHATE + DIACYLGLYCEROL.
-1- SIMILARITY: DOMAINS X AND Y ARE CONSERVED IN DIFFERENT FORMS OF
PLC AND ARE ESSENTIAL FOR CATALYTIC ACTIVITY.
-1- SIMILARITY: CONTAINS 1 C2 DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: Y13975; CAA74308.1; -
CC InterPro: IPR000008; C2;
CC InterPro: IPR001192; PL_PLC;
CC InterPro: IPR000909; PL_PLC_X;
CC InterPro: IPR001711; PL_PLC_Y;
CC Pfam: PF00388; PL_PLC-X; 2;
CC Pfam: PF00387; PL_PLC-Y; 1;
CC PRINTS: PR00390; PHPLIPASEC;
CC PRODOM: PD001202; PL_PLC_Y; 1;
CC SMART: SM00239; C2; 1;
CC SMART: SM00148; PLCXc; 1;
CC SMART: SM00149; PLCYc; 1;
CC PROSITE: PSS0004; C2_DOMAIN_2; 1;
CC PROSITE: PSS0007; PL_PLC_X_DOMAIN; 1;
CC PROSITE: PSS0008; PL_PLC_Y_DOMAIN; 1;
CC KEGG: Hydrolyase; Lipid degradation; Transducer.
CC FT DOMAIN 566 726 DOMAIN X;
CC FT DOMAIN 794 912 DOMAIN Y;
CC FT DOMAIN 938 1066 C2 DOMAIN;
CC FT ACT_SITE 579 579 BY SIMILARITY;
CC FT ACT_SITE 642 642 BY SIMILARITY;
CC FT DOMAIN 90 102 POLY-SER;
CC FT DOMAIN 239 243 POLY-THR;
CC FT DOMAIN 517 520 POLY-ASP;
CC FT DOMAIN 619 623 POLY-THR;
CC FT DOMAIN 629 632 POLY-ASP;
CC FT DOMAIN 743 760 POLY-THR;
CC FT DOMAIN 918 921 POLY-SER;
CC SO SEQUENCE 1099 AA; 124591 MW; D54D687D53A2829B CRC64;

Query Match 81.88; Score 27; DB 1; Length 1099;
Best Local Similarity 66.78; Pred. No. 3.2e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 FKXMR 6
DB 229 FKFSWR 234

```

```

"the gene encoding topoisomerase II from Plasmodium falciparum."
RL Nucleic Acids Res 22:2547-2551(1994).
CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
CC MAKES DOUBLE-STRAND BREAKS.
CC -1- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
CC OF DOUBLE-STRANDED DNA.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
CC RELAX ONLY NEGATIVE SUPERCOILS.
CC -1- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: X79345; -; NOT_ANNOTATED_CDS;
CC HSSP: P06786; 18GW;
CC InterPro: IPR000947; CBFA_NFYB;
CC InterPro: IPR001241; DNA_Topoisom;
CC InterPro: IPR002205; DNA_Topoisom;
CC InterPro: IPR003594; HATPase_C;
CC Pfam: PF00204; DNA_Topoisom; 1;
CC Pfam: PF00521; DNA_Topoisom; 1;
CC Pfam: PF02518; HATPase_C; 1;
CC PRINTS: PR00418; TP12FAMILY;
CC PRINTS: PR00615; CCANTSUBUNTA;
CC PRODOM: PD000616; DNA_Topoisom; 1;
CC SMART: SM00433; TOP2c; 1;
CC SMART: SM00434; TOP2c; 1;
CC PROSITE: PSS00177; TOPOISOMERASE_II; 1;
CC KEGG: Isomerase; Topoisomerase; DNA-binding; ATP-binding; Nuclear protein.
CC FT NP_BIND 144 149 ATP (POTENTIAL);
CC FT ACT_SITE 830 830 DNA_CLEAVAGE (BY SIMILARITY);
CC FT DOMAIN 271 281 POLY-ASN;
CC FT DOMAIN 308 316 POLY-ASN;
CC FT DOMAIN 1089 1093 POLY-LYS;
CC FT DOMAIN 1227 1234 POLY-LYS;
CC SO SEQUENCE 1398 AA; 161029 MW; BAAD7BEE88FE5B99 CRC64;

Query Match 81.88; Score 27; DB 1; Length 1398;
Best Local Similarity 80.08; Pred. No. 4e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 FKXMR 6
DB 173 FKMTW 177

```

```

RESULT 12
TOP2_PLAFK STANDARD; PRT; 1398 AA.
ID TOP2_PLAFK
AC P41001;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA TOPOISOMERASE II (EC 5.99.1.3).
GN TOP2.
OS Plasmodium falciparum (isolate KI / Thailand).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5839;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=94316496; PubMed=8041616;
RA Cheesman S., McAliese S., Goman M., Johnson D., Horrocks P.,
RA Ridley R.G., Klibey B.J.;

```

```

RESULT 13
YB51_ARCFU STANDARD; PRT; 108 AA.
ID YB51_ARCFU
AC O29114;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN AF1151.
GN AF1151.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN 11
RP SEQUENCE FROM N.A.
RX STRAIN=VC-16 / DSM 4304 / ATCC 49558;

```

```

RX MEDLINE-98049343: PubMed-9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.D., Kellavag A.R., Graham D.E., Kyrtides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.C., Gill S.,
RA Kirschner E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Cocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001024: AAB90099.1: -
DR TIGR: AF151: -
KW Hypothetical protein: Complete proteome.
SQ SEQUENCE 108 AA: 12829 MW: 94067725E39FA68 CRC64:

Query Match 78.8% Score 26: DB 1: Length 108:
Best Local Similarity 66.7% Pred. No. 59:
Matches 4: Conservative 0: Mismatches 2: Indels 0: Gaps 0:

OY 1 FKMWR 6
   1 1 1
Db 68 FRTAMR 73

RESULT 14
PT54_SACBA STANDARD: PRT: 294 AA.
AC 013364:
DT 15-JUL-1998 (rel. 36, Created)
DT 15-JUL-1998 (rel. 36, Last sequence update)
DT 30-MAY-2000 (rel. 39, Last annotation update)
DE PET54 PROTEIN.
CN PET54.
OS Saccharomyces bayanus (Yeast) (Saccharomyces uvarum).
OC Eukaryota: Fungi: Ascomycota: Saccharomycotina: Saccharomycetes:
OX NCBI_TaxID=4931:
   1 1
RN SEQUENCE FROM N.A.
RC STRAIN=NRRL Y-12624:
RA Costanzo M.C., Fox T.D.;
RL Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.
CC -I- FUNCTION: ACTIVATOR OF SPECIFIC MITOCHONDRIAL MRNAs. PET54
CC IS INVOLVED IN THE EXCISION OF INTRON AT5-BETA FROM PRE-MRNA
CC FOR CYTOCHROME C OXIDASE I (COX1) AND PLAYS A ROLE IN PROMOTING
CC THE TRANSLATION OF COX3 (BY SIMILARITY).
CC -I- SUBCELLULAR LOCATION: MITOCHONDRIAL; INNER MEMBRANE ASSOCIATED (BY
CC SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF023874: AAB81269.1: -

```

```

IR InterPro: IPR000504: RM.
IR SMART: SM00361: RM_L1: 1.
KW Mitochondrion; Inner membrane; RNA-binding; mRNA processing;
KW translation regulation; Activator.
SQ SEQUENCE 294 AA: 34770 MW: 430F9D477EEF88FC CRC64:

Query Match 78.8% Score 26: DB 1: Length 294:
Best Local Similarity 80.0% Pred. No. 15e+02:
Matches 4: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

OY 1 FKMW 5
   1 1 1
Db 227 FKMW 231

RESULT 15
YEJB_ECOLI STANDARD: PRT: 364 AA.
AC P33914: P76448:
DT 01-FEB-1994 (rel. 28, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 20-AUG-2001 (rel. 40, Last annotation update)
DE HYPOTHEICAL ABC TRANSPORTER PERMEASE PROTEIN YEJB.
GN YEJB OR B2178 OR Z3437 OR ECS3070.
OS Escherichia coli, and
US Escherichia coli O157:H7.
OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
OC Escherichia.
OX NCBI_TaxID=562, 83334:
   1 1
RN SEQUENCE FROM N.A.
RC STRAIN=K12 / BHB600:
RA Richterich P., Lahey N., Gryan G., Jaehn L., Mintz L., Robison K.,
RA Church G.M.;
RL Submitted (OCT-1993) to the EMBL/Genbank/DBJ databases.
   1 1
RN SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655:
EX MEDLINE-97426617: PubMed=9278503:
RA Blatterner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RL "The complete genome sequence of Escherichia coli K-12.";
FL Science 277:1453-1474(1997).
   1 1
RN SEQUENCE FROM N.A.
RC STRAIN=K12:
RC MEDLINE-97251358: PubMed=9097040:
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
RA Saito N., Sampei G., Seki Y., Sivasubram S., Tagami H.,
RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiiuchi T.;
RL "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
   1 1
RN SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDU933 / ATCC 700927:
RC MEDLINE-21074935: PubMed-11206551:
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posada G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grodzick E.J., Davis N.W., Lim A., Dimantanta E.T., Potamousis K.,
RA Welch R.A., Blatterner F.R.;
RL "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
   1 1
RN SEQUENCE FROM N.A.

```

KC STRAIN-0157:H7 / RIND 0509952;
 RX MEDLINE-21156231: PubMed-11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.,
 RT Complete genome sequence of enterohemorrhagic Escherichia coli
 RT 0157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 CC -!- FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT
 CC SYSTEM. PROBABLY RESPONSIBLE FOR THE TRANSLLOCATION OF THE
 CC SUBSTRATE ACROSS THE MEMBRANE.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
 CC (POTENTIAL).
 CC -!- SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER BINDING-
 CC PROTEIN-DEPENDENT TRANSPORT SYSTEMS. BELONGS TO THE OPBPC
 CC SUBFAMILY.
 CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
 CC FRAMESHIFT IN POSITION 126.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: U000008: AAA16376.1: ALT_FRAME.
 CC EMBL: AE000307: AAC75239.1: -.
 CC EMBL: D90849: BAA15986.1: -.
 CC DR EMBL: AE005450: AAG57316.1: -.
 CC DR EMBL: AP002560: BAB36493.1: -.
 CC DR Ecocore: EC12038: Ye3B.
 CC DR InterPro: IPR000515: BPD_transp.
 CC DR Pfam: PF00528: BPD_transp. 1.
 CC DR PROSITE: PS00402: BPD_TRANSF_INN_MEMBR: FALSE_NEG.
 CC KW Hypothetical protein; Transp; Transmembrane; Inner membrane;
 CC Complete proteome.
 CC MW TRANSMEM 9 29 POTENTIAL.
 CC FT TRANSMEM 38 58 POTENTIAL.
 CC FT TRANSMEM 136 156 POTENTIAL.
 CC FT TRANSMEM 173 193 POTENTIAL.
 CC FT TRANSMEM 225 245 POTENTIAL.
 CC FT TRANSMEM 285 305 POTENTIAL.
 CC FT TRANSMEM 327 347 POTENTIAL.
 CC SO SEQUENCE 364 AA: 40360 MW: 03AA6A7EC1F808AC CRC64;

Query Match 78.8%; Score 26; DB 1; Length 364;
 Best local Similarity 80.0%; Pred. No. 1.8e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 EKMXW 5
 1111
 Db 104 EKMLW 108

Search completed: February 27, 2002, 11:42:50
 Job time: 549 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:50:16 ; Search time 281.76 seconds
(without alignments)
3.115 Million cell updates/sec

Title: US-09-446-109A-20

Perfect score: 1.33

Sequence: 1 FKXMR 6

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPRMBL_17: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phage: *
10: sp_plant: *
11: sp_fodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	93.9	75	11 063050	063050 rattus norv
2	31	93.9	641	2 P70864	P70864 bartonella
3	31	93.9	989	11 063052	063052 rattus norv
4	29	87.9	353	5 09NGC1	09NGC1 drosophila
5	29	87.9	353	5 09NGC1	09NGC1 drosophila
6	29	87.9	353	5 09NGC9	09NGC9 drosophila
7	29	87.9	353	5 09NGA9	09NGA9 drosophila
8	29	87.9	353	5 09NGT5	09NGT5 drosophila
9	29	87.9	353	5 09NGT4	09NGT4 drosophila
10	29	87.9	784	5 09VC05	09VC05 drosophila
11	28	84.8	122	2 09A9K4	09A9K4 caulobacter
12	28	84.8	139	1 059213	059213 pyrococcus
13	28	84.8	207	11 09C201	09C201 mus musculu
14	28	84.8	289	2 09K704	09K704 bacillus ha
15	28	84.8	329	2 P96645	P96645 bacillus su
16	28	84.8	336	11 09W069	09W069 mus musculu
17	28	84.8	376	4 09Y2W5	09Y2W5 homo sapien
18	28	84.8	445	11 09DC12	09DC12 mus musculu
19	28	84.8	459	3 Q01446	Q01446 neustria hae

20	28	84.8	466	10 09ZRS1	09ZRS1 rictinus com
21	28	84.8	646	2 007625	007625 bacillus su
22	28	84.8	985	6 09XSW7	09XSW7 sus scrofa
23	28	84.8	1036	6 09TUN8	09TUN8 canis famli
24	28	84.8	1036	6 09TUN4	09TUN4 canis famli
25	28	84.8	1093	4 09P2E7	09P2E7 homo sapien
26	28	84.8	1141	11 09Z0X4	09Z0X4 mus musculu
27	28	84.8	1252	5 09Y0D0	09Y0D0 hydra atten
28	28	84.8	1412	10 064612	064612 arabidopsis
29	27	81.8	76	2 09JTE8	09JTE8 neisseria m
30	27	81.8	94	2 09JS37	09JS37 neisseria m
31	27	81.8	125	10 080923	080923 arabidopsis
32	27	81.8	177	10 09LSH6	09LSH6 arabidopsis
33	27	81.8	204	10 09FME5	09FME5 arabidopsis
34	27	81.8	221	2 09X6L2	09X6L2 klebsiella
35	27	81.8	280	5 045477	045477 caenorhabdi
36	27	81.8	297	4 043551	043551 homo sapien
37	27	81.8	305	4 043550	043550 homo sapien
38	27	81.8	360	2 09K0Y6	09K0Y6 neisseria m
39	27	81.8	360	2 09JSZ3	09JSZ3 neisseria m
40	27	81.8	417	5 017729	017729 caenorhabdi
41	27	81.8	434	10 09EXC1	09EXC1 arabidopsis
42	27	81.8	465	5 09GYP0	09GYP0 caenorhabdi
43	27	81.8	492	4 014926	014926 homo sapien
44	27	81.8	492	6 018728	018728 bos taurus
45	27	81.8	539	4 013971	013971 homo sapien

ALIGNMENTS

RESULT 1
C63050
ID 063050 PRELIMINARY; PRT: 75 AA.
AC 063050:
DT 01-NOV-1996 (TRENBLREL. 01, Created)
DT 01-JAN-1999 (TRENBLREL. 09, last sequence update)
DT 01-JAN-1999 (TRENBLREL. 09, last annotation update)
FT 01-JAN-1999 (TRENBLREL. 09, last annotation update)
LE Rattus norvegicus (Rat).
CS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
GX NCBI_TaxID=10116;
RN [1]
RP .SEQUENCE FROM N.A.
EX MEDLINE=85270375; PubMed=3860811;
FA Lusis A.J., West R., Mehrlan M., Reuben M.A., LeBoeuf R.C.,
RA Kaptein J.S., Johnson D.F., Schumaker V.N., Yuhasz M.P., Scholtz M.C.,
EA Elverson J.;
RT "Cloning and expression of apolipoprotein B, the major protein of low
RT and very low density lipoproteins."
RL Proc. Natl. Acad. Sci. U.S.A. 82:4597-4601(1985).
LR EMBL: M11227; AAA40752.1; -
KW Lipoprotein.
FT NON_TER 1 1
FT NON_TER 75 75
FT SEQUENCE 75 AA: 8787 MW: F8B789B42BC5E489 CRC64;
Query Match 93.9% Score 31: DB 11: Length 75:
Best Local Similarity 83.3% Pred. No. 22:
Matches 5: Conservative 0: Mismatches 1: Indels 0: Gaps 0:
CY 1 FKXMR 6
DB 3 FKXMR 8
RESULT 2
ID P70864 PRELIMINARY; PRT: 641 AA.
AC P70864:
DT 01-FEB-1997 (TRENBLREL. 02, Created)

DT 01-FEB-1997 (Tremblrel. 02, last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
 DE TXPA (FRAGMENT).
 GN TXPA.
 OS Bartonella bacilliformis.
 CC Bacteria: Proteobacteria: alpha subdivision: Rhizobiaceae group:
 CC Bartonellaceae: Bartonella.
 CC NCBI_TaxID=774;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA upetiace E., Inler G.M.;
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
 CC -i- SIMILARITY: TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC
 TRANSPORTERS).
 CC EMBL: U68242; AAB09036.1; -
 DR HSSP: P13569; INBD.
 DR InterPro: IPR003593; AAA.
 DR InterPro: IPR001140; ABC_transporter_tmem.
 DR InterPro: IPR003439; ABC_transporter.
 DR InterPro: IPR001687; ATP_GTP_A.
 DR Pfam: PF00664; ABC_membrane; 1.
 DR Pfam: PF00005; ABC_tran; 1.
 DR SMART: SM00382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 DR ATP-binding; Transport.
 FT NON_TER 1
 SO SEQUENCE 641 AA: 71542 MW: 0E28FAA9A1828EFB CRC64;

Query Match 93.9%; Score 31; DB 2; Length 641;
 Best Local Similarity 83.3%; Pred. No. 1.7e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKXMR 6
 DB 198 FKXMR 203

RESULT 3
 Q63052 PRELIMINARY: PRT; 989 AA.
 AC Q63052;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
 DT 01-NOV-1998 (Tremblrel. 08, last annotation update)
 DE APOLIPROTEIN B (FRAGMENT).
 GN APOB.
 OS Rattus norvegicus (Rat).
 CC Eukaryota: Metazoa: Chordata: Vertebrata: Euteleostomi:
 CC Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Rattus.
 CC NCBI_TaxID=10116;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RX MEDLINE=89176719; PubMed=3235917;
 RA Reuben M.A., Svenson K.L., Doolittle M.H., Johnson D.F., Lusis A.J.,
 Elovson J.;
 RT *Biosynthetic relationships between three rat apolipoprotein B
 peptides.*;
 RL J. Lipid Res. 29:1337-1347(1988).
 DR EMBL: M27440; AAA74690.1; -
 KW Lipoprotein.
 FT NON_TER 1
 SO SEQUENCE 989 AA: 113305 MW: 21E9D5149100F9E4 CRC64;

Query Match 93.9%; Score 31; DB 11; Length 989;
 Best Local Similarity 83.3%; Pred. No. 2.5e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKXMR 6
 DB 488 FKXMR 493

RESULT 4
 Q9NGC1 PRELIMINARY: PRT; 353 AA.
 ID Q9NGC1;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
 DE OXYSTEROL BINDING PROTEIN (FRAGMENT).
 GN OSBP.
 OS Drosophila simulans (Fruit fly).
 CC Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:
 CC Pterygota: Neoptera: Endopterygota: Diptera: Brachycera: Muscomorpha:
 CC Ephydroidea: Drosophilidae: Drosophila.
 CC NCBI_TaxID=7240;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA STRAIN=SIM2;
 RX MEDLINE=20283933; PubMed=10823947;
 RA Begun D.J., Whitely P.;
 RT *Reduced X-linked nucleotide polymorphism in Drosophila simulans.*;
 RL Proc. Natl. Acad. Sci. U.S.A. 97:5960-5965(2000).
 DR EMBL: AF252730; AAF68513.1; -
 DR FlyBase: FBgn0041668; Dsim\Osbp.
 DR InterPro: IPR000648; Oysterol_BP.
 DR Pfam: PF01237; Oysterol_BP; 1.
 DR PROSITE: PS01013; OSBP; 1.
 FT NON_TER 1
 SO SEQUENCE 353 AA: 41011 MW: 3FF49E1A2336F0F9 CRC64;

Query Match 87.9%; Score 29; DB 5; Length 353;
 Best Local Similarity 66.7%; Pred. No. 2.4e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKXMR 6
 DB 228 FKXMR 233

RESULT 5
 Q9NGC0 PRELIMINARY: PRT; 353 AA.
 ID Q9NGC0;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, last annotation update)
 DE OXYSTEROL BINDING PROTEIN (FRAGMENT).
 GN OSBP.
 OS Drosophila simulans (Fruit fly).
 CC Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:
 CC Pterygota: Neoptera: Endopterygota: Diptera: Brachycera: Muscomorpha:
 CC Ephydroidea: Drosophilidae: Drosophila.
 CC NCBI_TaxID=7240;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=SIM7;
 RX MEDLINE=20283933; PubMed=10823947;
 RA Begun D.J., Whitely P.;
 RT *Reduced X-linked nucleotide polymorphism in Drosophila simulans.*;
 RL Proc. Natl. Acad. Sci. U.S.A. 97:5960-5965(2000).
 DR EMBL: AF252735; AAF68518.1; -
 DR FlyBase: FBgn0041668; Dsim\Osbp.
 DR InterPro: IPR000648; Oysterol_BP.
 DR Pfam: PF01237; Oysterol_BP; 1.
 DR PROSITE: PS01013; OSBP; 1.
 FT NON_TER 1
 SO SEQUENCE 353 AA: 41072 MW: 0497E71CF85A2CA3 CRC64;

Query Match 87.9%: Score 29; DB 5; Length 353;
 Best Local Similarity 66.7%: Pred. No. 2.4e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKMWR 6
 11: 11
 DB 228 FKLAMR 233

RESULT 6

O9NCA9 PRELIMINARY: PRT: 353 AA.

ID O9NCA9: 01-OCT-2000 (TREMBLrel. 15, Created)
 AC O9NCA9: 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 17, Last annotation update)
 DE OXYSTEROL BINDING PROTEIN (FRAGMENT).

OS Drosophila yakuba (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7245;

RP SEQUENCE FROM N.A.
 RX MEDLINE=20283933; PubMed=10823947;
 RA Begun D.J., Whitley P.;

RT "Reduced X-linked nucleotide polymorphism in Drosophila simulans."
 RL Proc. Natl. Acad. Sci. U.S.A. 97:5960-5965(2000).

DR EMBL: AF252732; AAF68516.1; -
 DR FlyBase: FBgn0041644; DyakOsbp.
 DR InterPro: IPR000648; Oxysterol_BP.
 DR Pfam: PF01237; Oxysterol_BP. 1.

DR PROSITE: PS01013; OSBP. 1.
 FT NON_TER 1
 FT NON_TER 353

SO SEQUENCE 353 AA; 40975 MW; F6C52ECA897B5DAD CRC64;

Query Match 87.9%: Score 29; DB 5; Length 353;
 Best Local Similarity 66.7%: Pred. No. 2.4e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKMWR 6
 11: 11
 DB 228 FKLAMR 233

RESULT 7

O9N6T5 PRELIMINARY: PRT: 353 AA.

ID O9N6T5: 01-OCT-2000 (TREMBLrel. 15, Created)
 AC O9N6T5: 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 17, Last annotation update)
 DE OXYSTEROL BINDING PROTEIN (FRAGMENT).

OS Drosophila simulans (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7240;

RP SEQUENCE FROM N.A.
 RX MEDLINE=20283933; PubMed=10823947;
 RA Begun D.J., Whitley P.;

RT "Reduced X-linked nucleotide polymorphism in Drosophila simulans."
 RL Proc. Natl. Acad. Sci. U.S.A. 97:5960-5965(2000).
 DR EMBL: AF252733; AAF68516.1; -
 DR EMBL: AF252729; AAF68512.1; -
 DR FlyBase: FBgn0041668; DsimOsbp.
 DR InterPro: IPR000648; Oxysterol_BP.

DR Pfam: PF01237; Oxysterol_BP. 1.
 DR PROSITE: PS01013; OSBP. 1.
 FT NON_TER 1
 FT NON_TER 353
 SO SEQUENCE 353 AA; 41045 MW; 049928CCF85ED073 CRC64;

Query Match 87.9%: Score 29; DB 5; Length 353;
 Best Local Similarity 66.7%: Pred. No. 2.4e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKMWR 6
 11: 11
 DB 228 FKLAMR 233

RESULT 8

O9N6T4 PRELIMINARY: PRT: 353 AA.

ID O9N6T4: 01-OCT-2000 (TREMBLrel. 15, Created)
 AC O9N6T4: 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 17, Last annotation update)
 DE OXYSTEROL BINDING PROTEIN (FRAGMENT).

OS Drosophila simulans (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7240;

RP SEQUENCE FROM N.A.
 RX MEDLINE=20283933; PubMed=10823947;
 RA Begun D.J., Whitley P.;

RT "Reduced X-linked nucleotide polymorphism in Drosophila simulans."
 RL Proc. Natl. Acad. Sci. U.S.A. 97:5960-5965(2000).

DR EMBL: AF252732; AAF68515.1; -
 DR EMBL: AF252734; AAF68517.1; -
 DR FlyBase: FBgn0041668; DsimOsbp.
 DR InterPro: IPR000648; Oxysterol_BP.
 DR Pfam: PF01237; Oxysterol_BP. 1.

DR PROSITE: PS01013; OSBP. 1.
 FT NON_TER 1
 FT NON_TER 353

SO SEQUENCE 353 AA; 40985 MW; 049928D652FEC173 CRC64;

Query Match 87.9%: Score 29; DB 5; Length 353;
 Best Local Similarity 66.7%: Pred. No. 2.4e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKMWR 6
 11: 11
 DB 228 FKLAMR 233

RESULT 9

O18397 PRELIMINARY: PRT: 784 AA.

ID O18397: 01-JAN-1998 (TREMBLrel. 05, Created)
 AC O18397: 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE OXYSTEROL BINDING PROTEIN HOMOLOGUE.

OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;

```

RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=98146266; PubMed=9473651;
RA      Alpey L.S., Jimenez J., Glover D.M.;
RT      "A Drosophila homologue of oxysterol binding protein
RT      (OSBP)--implications for the role of OSBP.";
RL      Blochin. Biophys. Acta 1395:159-164(1998).
DR      EMBL: Y13951; CAA74289.1; -
DR      Flybase: FBgn0020626; Osbp.
DR      InterPro: IPR000648; Oxysterol_BP.
DR      InterPro: IPR001849; PH.
DR      Pfam: PF01237; Oxysterol_BP. 1.
DR      Pfam: PF00169; PH. 1.
DR      SMART: SM00233; PH. 1.
DR      PROSITE: PS01013; OSBP. 1.
DR      PROSITE: PS50003; PH_DOMAIN. 1.
SQ      .SEQUENCE 784 AA; 89234 MW; BACE6DB18B45A37D CRC64;

Query Match          87.9%; Score 29; DB 5; Length 784;
Best Local Similarity 66.7%; Pred. No. 5e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 FRKXMR 6
DB      654 FKLAWR 659

RESULT  10
Q9VCO5  PRELIMINARY; PRT; 784 AA.
ID      01-MAY-2000 (TREMBlrel. 13, Created)
DT      01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT      01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE      OSBP PROTEIN.
CN      OSBP OR CG6708.
OS      Drosophila melanogaster (Fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC      Ephydroidea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC      Ephydroidea; Drosophilidae; Drosophila.
OX      NCBI_TaxID=7227;
[1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=BERKELEY.
RX      MEDLINE=20196006; PubMed=10731132;
RA      Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA      Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA      George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA      Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA      Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA      Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA      Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA      Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA      Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolintsov S.,
RA      Botkova D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,
RA      Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA      Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA      de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA      Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA      Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA      Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA      Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA      Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA      Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA      Jaitai M., Kalush K., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA      Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA      Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA      Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA      Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA      Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA      Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclob J.M.,
RA      Palazzolo M., Piltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

```

```

RA      Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA      Shue B.C., Siden-Klimos I., Simpson M., Skupski M.P., Smith T.,
RA      Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA      Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA      Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA      Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA      Ye J., Yeh R.-F., Zaveli J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA      Zheng X.H., Zhong E.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA      Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT      "The genome sequence of Drosophila melanogaster.";
RL      Science 287:2185-2195(2000).
DR      EMBL: AE003749; AAF56371.1; -
DR      Flybase: FBgn0020626; Osbp.
DR      InterPro: IPR000648; Oxysterol_BP.
DR      InterPro: IPR001849; PH.
DR      Pfam: PF01237; Oxysterol_BP. 1.
DR      Pfam: PF00169; PH. 1.
DR      SMART: SM00233; PH. 1.
DR      PROSITE: PS01013; OSBP. 1.
DR      PROSITE: PS50003; PH_DOMAIN. 1.
SQ      .SEQUENCE 784 AA; 89324 MW; E11429C115C0C80D CRC64;

Query Match          87.9%; Score 29; DB 5; Length 784;
Best Local Similarity 66.7%; Pred. No. 5e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 FRKXMR 6
DB      654 FKLAWR 659

RESULT  11
Q9A9K4  PRELIMINARY; PRT; 122 AA.
ID      01-JUN-2001 (TREMBlrel. 17, Created)
DT      01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT      01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE      HYPOTHETICAL PROTEIN CC0976.
CN      CC0976.
OS      Caulobacter crescentus.
OC      Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC      Caulobacter.
OX      NCBI_TaxID=69394;
[1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=21173698; PubMed=11259647;
RA      Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA      Eisen J., Heidelberg J.F., Alley M.R.K., Onta N., Maddock J.R.,
RA      Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA      Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA      Kolonay J.F., Smit J., Craven M.B., Khouri H., Shelly J., Berry K.,
RA      Ullendack T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA      Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT      "Complete genome sequence of Caulobacter crescentus.";
RL      Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR      EMBL: AE005775; AAK22960.1; -
DR      TIGR: CC0976; -
KW      Hypothetical protein; Complete proteome.
SQ      .SEQUENCE 122 AA; 13642 MW; 389E58C4C4DE2269 CRC64;

Query Match          84.8%; Score 28; DB 2; Length 122;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 FRKXMR 6
DB      21 FKLWMR 26

RESULT  12

```



```

059213
ID 059213 PRELIMINARY: PRT: 139 AA.
AC 059213
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE HYPOTHETICAL 16.2 KDA PROTEIN PH1559.
CN PH1559.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Halkawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
RL DNA Res. 5:55-76(1998)
DR EMBL: AP000006; BAA30671.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 139 AA: 16229 MW: 5A23FA7D56875382 CRC64;

Query Match 84.8%; Score 28; DB 1; Length 139;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKMNR 6
DB 68 FKLNR 73

RESULT 13
09C201 PRELIMINARY: PRT: 207 AA.
ID 09C201
AC 09C201
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 2700031G06RIK PROTEIN.
CN 2700031G06RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO.
RX MEDLINE=21083660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Alizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Cojocari T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schiml L.M., Staubl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Norcone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,
RA Wyszynski A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;

```

```

RT "Functional annotation of a full-length mouse cDNA collection.";
PL Nature 409:685-690(2001);
DR EMBL: AK012307; BAB28153.1;
DR MGD: MGI:1914443; 2700031G06RIK.
DR InterPro: IPR000092; PolyPrenyl_synth.
DR PROSITE: PS00723; POLYPRENYL_SYNTHTF.1;
SQ SEQUENCE 207 AA: 23200 MW: 42A34D18D0A5723F CRC64;

Query Match 84.8%; Score 28; DB 1; Length 207;
Best Local Similarity 66.7%; Pred. No. 2.3e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKMNR 6
DB 17 FKLNR 22

RESULT 14
09K704 PRELIMINARY: PRT: 289 AA.
ID 09K704
AC 09K704
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE FORMYLDETRAHYDROFLATE DEFORMYLASE.
CN B113265.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuniara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL: AP001518; BAB06984.1;
DR InterPro: IPR002912; ACT.
DR InterPro: IPR002376; formyl-transf.
DR Pfam: PF01842; ACT.1.
DR Pfam: PF00551; formyl-transf.1.
KW Complete proteome.
SQ SEQUENCE 289 AA: 33404 MW: 7E7D7EF58D300806 CRC64;

Query Match 84.8%; Score 28; DB 2; Length 289;
Best Local Similarity 66.7%; Pred. No. 3.1e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKMNR 6
DB 80 YKMNR 85

RESULT 15
P96645 PRELIMINARY: PRT: 329 AA.
ID P96645
AC P96645
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE YDHI PROTEIN.
CN YDHI.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-168:
RX MEDLINE-98000887: PubMed-9341680:
RA Beloin C., Ayora S., Exley R., Hirschbein L., Ogasawara N.,
RA Kasahara Y., Alonso J.C., Le Hegarat F.,
RT "Characterization of an lrp-like (lrpC) gene from Bacillus subtilis."
RL Mol. Gen. Genet. 256:63-71(1997).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-168:
RX MEDLINE-98044033: PubMed-9384377:
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo J., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Enlian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Coffeau A., Gollightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogilwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchilyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambuit R., Wedler E., Wedler H., Weltzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN-168:
RX Kunst F., Ogasawara N., Yoshikawa H., Danchin A.,
RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL: AB001488: BAA19334.1: -
DR EMBL: Z99106: CAB12304.1: -
DR InterPro: IPR000064: NLPC_P60.
DR InterPro: IPR000189: SLT.
DR Pfam: PF00877: NLPC_P60: 1.
DR Pfam: PF01464: SLT: 1.
DR Complete proteome.
SQ SEQUENCE 329 AA: 36546 MW: BLAD04875F5C38E CRC64:

```

```

Query Match      84.8%: Score 28: DB 2: Length 329:
Best Local Similarity 66.7%: Pred. No. 3.5e+02:
Matches 4: Conservative 1: Mismatches 1: Indels 0: Gaps 0:

```

```

OY 1 FKMxMR 6
DB 172 FKMxMR 177

```

Search completed: February 27, 2002, 11:50:17
Job time: 996 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:36:09 ; Search time 132.19 Seconds
(without aliases)

Title: US-09-446-109A-20

Perfect score: 33
Sequence: 1 FKMXXR 6

1 EKMXWR 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```
Database :
Issued_Patents_AA :
1 : /cgn2_6/prodata/2/1aa/5A.CONB.pep.*
2 : /cgn2_6/prodata/2/1aa/5B.CONB.pep.*
3 : /cgn2_6/prodata/2/1aa/6A.CONB.pep.*
4 : /cgn2_6/prodata/2/1aa/6B.CONB.pep.*
5 : /cgn2_6/prodata/2/1aa/PCUTS.CONB.pep.*
6 : /cgn2_6/prodata/2/1aa/backfiles1.pep.*
```

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

No.	Score	Query Match	Length	DB	ID	Description
1	27	81.8	566	1	US-08-428-415-4	Sequence 4, Appl1
2	27	81.8	566	1	US-08-319-685-4	Sequence 4, Appl1
3	27	81.8	566	3	US-08-854-029-4	Sequence 4, Appl1
4	27	81.8	566	3	US-08-848-810-2	Sequence 2, Appl1
5	27	81.8	566	4	US-08-428-762-4	Sequence 4, Appl1
6	26	78.8	37	1	US-08-665-5438-4	Sequence 4, Appl1
7	26	78.8	50	2	US-08-248-839C-117	Sequence 117, Appl1
8	26	78.8	1144	2	US-08-147-812-5	Sequence 5, Appl1
9	26	78.8	1144	2	US-08-319-866-12	Sequence 12, Appl1
10	26	78.8	1144	4	US-09-123-708-2	Sequence 2, Appl1
11	26	78.8	1144	4	US-09-123-624-2	Sequence 2, Appl1
12	25	75.8	14	1	US-08-383-753-11	Sequence 31, Appl1
13	25	75.8	14	2	US-08-586-712-11	Sequence 31, Appl1
14	25	75.8	14	2	US-08-959-512-11	Sequence 31, Appl1
15	25	75.8	14	4	US-09-512-983-11	Sequence 31, Appl1
16	25	75.8	15	1	US-08-383-753-35	Sequence 35, Appl1
17	25	75.8	15	1	US-08-383-753-40	Sequence 40, Appl1
18	25	75.8	15	1	US-08-383-753-46	Sequence 46, Appl1
19	25	75.8	15	2	US-08-586-772-35	Sequence 35, Appl1
20	25	75.8	15	2	US-08-586-772-40	Sequence 40, Appl1
21	25	75.8	15	2	US-08-586-772-46	Sequence 46, Appl1
22	25	75.8	15	2	US-08-959-512-35	Sequence 35, Appl1
23	25	75.8	15	2	US-08-959-512-40	Sequence 40, Appl1
24	25	75.8	15	2	US-08-959-512-46	Sequence 46, Appl1
25	25	75.8	15	4	US-09-512-983-35	Sequence 35, Appl1
26	25	75.8	15	4	US-09-512-983-40	Sequence 40, Appl1
27	25	75.8	15	4	US-09-512-983-46	Sequence 46, Appl1

28	25	75.8	20	3	US-08-504-538A-11	Sequence 11, App1
29	25	75.8	20	5	PCT-US95-09307-11	Sequence 11, App1
30	25	75.8	23	1	US-08-383-753-74	Sequence 74, App1
31	25	75.8	23	1	US-08-383-753-80	Sequence 80, App1
32	25	75.8	23	1	US-08-383-753-85	Sequence 85, App1
33	25	75.8	23	1	US-08-586-772-74	Sequence 74, App1
34	25	75.8	23	2	US-08-586-772-80	Sequence 80, App1
35	25	75.8	23	2	US-08-586-772-85	Sequence 85, App1
36	25	75.8	23	2	US-08-959-512-74	Sequence 74, App1
37	25	75.8	23	2	US-08-959-512-80	Sequence 80, App1
38	25	75.8	23	2	US-08-959-512-85	Sequence 85, App1
39	25	75.8	23	4	US-09-512-983-74	Sequence 74, App1
40	25	75.8	23	4	US-09-512-983-80	Sequence 80, App1
41	25	75.8	23	4	US-09-512-983-85	Sequence 85, App1
42	25	75.8	45	2	US-08-248-839C-158	Sequence 158, App1
43	25	75.8	45	2	US-08-889-425-2	Sequence 2, App1
44	25	75.8	45	2	US-08-764-870-16	Sequence 16, App1
45	25	75.8	45	2	US-08-980-115-16	Sequence 16, App1

ALIGNMENTS

RESULT 1
 HS-08-428-415-4
 Sequence 4, Application US/08428415
 Patent No. 5756335
 GENERAL INFORMATION:
 APPLICANT: Cold Spring Harbor Laboratory
 TITLE OF INVENTION: No. 5756335el Human cd25 Genes, Encoded
 TITLE OF INVENTION: Products and Uses Thereof
 NUMBER OF SEQUENCES: 31
 CORRESPONDENCE ADDRESS:
 ADDRESSER: LAHIVE & COCKFIELD
 STREET: 60 State Street
 CITY: Boston
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: ASCII(text)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/428,415
 FILING DATE: 24 April 1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Vincent, Matthew P.
 REGISTRATION NUMBER: 36,709
 REFERENCE/DOCKET NUMBER: MII-019CP3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-227-7400
 TELEFAX: 617-227-5941
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 566 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-428-415-4

Query match 81.8%; Score 27; DB 1; Length 566;

```

cy      1 EKMXWR 6
        ||| |
jb      185 EKMPWK 190

```

RESULT 2
US-08-379-685-4
Sequence 4, Application US/08379685
Patent No. 5770423
GENERAL INFORMATION:
APPLICANT: Cold Spring Harbor Laboratory
TITLE OF INVENTION: No. 5770423el Human cdc25 Genes, Encoded
TITLE OF INVENTION: Products and Uses Thereof
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII(text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/379,685
FILING DATE: 26 January 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MII-019-DV
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-227-7400
TELEFAX: 617-227-5941
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 566 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-379-685-4

Query Match 81.8%; Score 27; DB 1; Length 566;
Best Local Similarity 66.7%; Pred. No. 5e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FRMXR 6
DB 185 FKMPWK 190

RESULT 3
US-08-854-029-4
Sequence 4, Application US/08854029
Patent No. 5994074
GENERAL INFORMATION:
APPLICANT: Cold Spring Harbor Laboratory
TITLE OF INVENTION: No. 5994074el Human cdc25 Genes, Encoded
TITLE OF INVENTION: Products and Uses Thereof
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HOAG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII(text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/854,029

FILING DATE: 2 MAY 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MII-019, 08
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1242
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 566 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-854-029-4

Query Match 81.8%; Score 27; DB 2; Length 566;
Best Local Similarity 66.7%; Pred. No. 5e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FRMXR 6
DB 185 FKMPWK 190

RESULT 4
US-08-848-810-2
Sequence 2, Application US/08848810
Patent No. 6074851
GENERAL INFORMATION:
APPLICANT: Deibel Jr., M. R.
APPLICANT: Yem, A. W.
TITLE OF INVENTION: Catalytic Macro Molecules Having DCD25B
TITLE OF INVENTION: Like Activity
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pharmacia & Upjohn Company
STREET: 301 Henrietta Street
CITY: Kalamazoo
STATE: MI
COUNTRY: USA
ZIP: 49001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/848,810
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Woolton, Thomas A.
REGISTRATION NUMBER: 35,004
TELECOMMUNICATION INFORMATION:
TELEPHONE: 616-833-7914
TELEFAX: 616-833-6897
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 566 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-848-810-2

Query Match 81.8%; Score 27; DB 3; Length 566;

Best Local Similarity 66.7%; Pred. No. 5e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKKXMR 6
1111:
Db 185 FKKMPWK 190

RESULT 5

US-08-428-762-4
Sequence 4, Application US/08428762
Patent No. 6251585
GENERAL INFORMATION:
APPLICANT: Cold Spring Harbor Laboratory
TITLE OF INVENTION: No. 6251585el Human cdc25 Genes, Encoded
TITLE OF INVENTION: Products and Uses Thereof
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII(text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/428,762
FILING DATE: 24 April 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MIT-019-DV
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-227-7400
TELEFAX: 617-227-5941
INFORMATION FOR SEQ. ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 566 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-428-762-4

Query Match 81.8%; Score 27; DB 4; Length 566;
Best Local Similarity 66.7%; Pred. No. 5e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKKXMR 6
1111:
Db 185 FKKMPWK 190

RESULT 6
US-08-665-543B-4
Sequence 4, Application US/0866543B
Patent No. 5734015
GENERAL INFORMATION:
APPLICANT: Shinnar, Ann
APPLICANT: Zaslloff, Michael A.
APPLICANT: Uzzell, Thomas
TITLE OF INVENTION: NEW FAMILY OF LINEAR ANTIMICROBIAL
TITLE OF INVENTION: PEPTIDES FROM HAGRISH INTESTINE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flinnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.

CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/665,543B
FILING DATE: 18-JUN-1996
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/000,325
FILING DATE: 19-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Barlow, Stacey A.
REGISTRATION NUMBER: 39,595
REFERENCE/DOCKET NUMBER: 05387, 0036-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ. ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-665-543B-4

Query Match 78.8%; Score 26; DB 1; Length 37;
Best Local Similarity 66.7%; Pred. No. 58;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FKKXMR 6
1111:
Db 3 FKKXMR 8

RESULT 7

US-08-248-839C-117
Sequence 117, Application US/08248839C
Patent No. 5843702
GENERAL INFORMATION:
APPLICANT: McConnell, David
APPLICANT: Devine, Kevin
APPLICANT: O'Kane, Charles
TITLE OF INVENTION: A Gene Expression System
NUMBER OF SEQUENCES: 185
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 58437020 No. 5843702disk of No. 5843702ch America, Inc.
STREET: 405 Lexington Avenue.
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/248,839C
FILING DATE: 23-MAY-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gregg, Valeta A.
REGISTRATION NUMBER: 35,127
REFERENCE/DOCKET NUMBER: 3614, 214-US
TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 117:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-08-248-839C-117

Query Match: 78.8% Score 26; DB 2; Length 50;
Best Local Similarity 66.7% Pred. No. 77;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Caps 0;

QY 1 FKXMR 6
1 1 1 1
DB 21 FSLMR 26

RESULT 8
US-08-147-812-5
Sequence 5, Application US/08147812
Patent No. 5766909
GENERAL INFORMATION:
APPLICANT: Xie, Qiao-wen
APPLICANT: Nathan, Carl F.
APPLICANT: Mumford, Richard A.
APPLICANT: Calaycay, Jimmy Ramos
TITLE OF INVENTION: DNA Encoding Inducible Nitric Oxide Synthase
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: 126 East Lincoln Avenue
CITY: Rahway
STATE: New Jersey
COUNTRY: USA
ZIP: 07065

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: Macintosh Centris650
OPERATING SYSTEM: Macintosh 7.0.1
SOFTWARE: Microsoft Word 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/147,812
FILING DATE: NO. 5766909 Available

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/841,641
FILING DATE: 02-FEB-1992

ATTORNEY/AGENT INFORMATION:
NAME: Wallen, John W III
REGISTRATION NUMBER: 35,403
REFERENCE/DOCKET NUMBER: 186581A
TELEPHONE: (908) 594-3905
TELEFAX: (908) 594-4720
TELEX: 138825

INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1144 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-147-812-5

Query Match: 78.8% Score 26; DB 1; Length 1144;
Best Local Similarity 80.0% Pred. No. 1.5e+03;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 2 KXMR 6
1 1 1 1

DB 185 KXMR 189

RESULT 9
US-08-319-866-12
Sequence 12, Application US/08319866
Patent No. 5929223
GENERAL INFORMATION:

APPLICANT: Tully, Timothy P.
APPLICANT: Yin, Jerry C.
APPLICANT: Regulski, Michael
TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF GENES
ASSOCIATED WITH LONG-TERM MEMORY
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millita Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/319,866
FILING DATE: 7-OCT-1994
CLASSIFICATION: 514

PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: CSHL94-03
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540

INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1144 amino acids
TYPE: amino acid
STRANDEDNESS:

TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-319-866-12

Query Match: 78.8% Score 26; DB 2; Length 1144;
Best Local Similarity 80.0% Pred. No. 1.5e+03;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 2 KXMR 6
1 1 1 1
DB 185 KXMR 189

RESULT 10
US-09-123-708-2
Sequence 2, Application US/09123708
Patent No. 6146887

GENERAL INFORMATION:
APPLICANT: SCHRAEDER, Juergen
APPLICANT: GOEDECKE, Axel

TITLE OF INVENTION: DNA EXPRESSION VECTORS FOR USE IN GENE THERAPEUTIC
TREATMENT OF VASCULAR DISORDERS
FILE REFERENCE: 511169-2003
CURRENT APPLICATION NUMBER: US/09/123,708
CURRENT FILING DATE: 1998-07-28
EARLIER APPLICATION NUMBER: 08/553,503
EARLIER FILING DATE: 1996-03-01

EARLIER APPLICATION NUMBER: P4411402.8
; EARLIER FILING DATE: 1994-03-31
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO: 2
; LENGTH: 1144
; TYPE: PRT
; ORGANISM: Cytomegalovirus
US-09-123-708-2

Query Match 78.8%; Score 26; DB 4; Length 1144;
Best Local Similarity 80.0%; Pred. No. 1.5e+03;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 KMXMR 6
DB 185 KMAWR 189

RESULT 11
US-09-123-624-2
; Sequence 2, Application US/09123624
; Patent No. 6149936
; GENERAL INFORMATION:
; APPLICANT: SCHRADER, Jurgen
; APPLICANT: GODECKE, Axel
; TITLE OF INVENTION: DNA EXPRESSION VECTORS FOR USE IN THE GENE THERAPEUTIC
; FILE REFERENCE: 511169-2004
; CURRENT APPLICATION NUMBER: US/09/123,624
; CURRENT FILING DATE: 1998-07-28
; PRIOR APPLICATION NUMBER: 08/553,503
; PRIOR FILING DATE: 1996-03-01
; PRIOR APPLICATION NUMBER: 4411402.8
; PRIOR FILING DATE: 1994-03-31
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO: 2
; LENGTH: 1144
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-123-624-2

Query Match 78.8%; Score 26; DB 4; Length 1144;
Best Local Similarity 80.0%; Pred. No. 1.5e+03;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 KMXMR 6
DB 185 KMAWR 189

RESULT 12
US-08-383-753-31
; Sequence 31, Application US/08383753
; Patent No. 5723584
; GENERAL INFORMATION:
; APPLICANT: Schatz, Peter J.
; TITLE OF INVENTION: Biotinylation of Proteins
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourlie and Crew
; STREET: One Market Plaza, Stewart Tower
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/383,753
FILING DATE: 03-FEB-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/099,991
FILING DATE: 30-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Smith, William M.
REGISTRATION NUMBER: 30,223
REFERENCE/DOCKET NUMBER: 1038.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-383-753-31

Query Match 75.8%; Score 25; DB 1; Length 14;
Best Local Similarity 80.0%; Pred. No. 36;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 KMXMR 6
DB 9 KMLMR 13

RESULT 13
US-08-586-772-31
; Sequence 31, Application US/08586772
; Patent No. 5874239
; GENERAL INFORMATION:
; APPLICANT: Schatz, Peter J.
; TITLE OF INVENTION: Biotinylation of Proteins
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourlie and Crew
; STREET: One Market Plaza, Stewart Tower
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/586,772
; FILING DATE: 03-FEB-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/099,991
; FILING DATE: 30-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, William M.
; REGISTRATION NUMBER: 30,223
; REFERENCE/DOCKET NUMBER: 1038.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-586-772-31

Query Match 75.8% Score 25; DB 2; Length 14;
Best Local Similarity 80.0% Pred. No. 36;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 KMXWR 6
11 11
DB 9 KMLMR 13

RESULT 14
US-08-959-512-31
Sequence 31, Application US/08959512
Patent No. 5932433
GENERAL INFORMATION:
APPLICANT: Schatz, Peter J.
TITLE OF INVENTION: Biocinylation of Proteins
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Kourile and Crew
STREET: One Market Plaza, Stewart Tower
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentia Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/959,512
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/383,753
FILING DATE: 03-FEB-1995
APPLICATION NUMBER: US 08/099,991
FILING DATE: 30-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Smith, William M.
REGISTRATION NUMBER: 30,223
REFERENCE/DOCKET NUMBER: 1038.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-959-512-31

Query Match 75.8% Score 25; DB 2; Length 14;
Best Local Similarity 80.0% Pred. No. 36;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 KMXWR 6
11 11
DB 9 KMLMR 13

RESULT 15
US-09-512-983-31
Sequence 31, Application US/09512983
Patent No. 6265552

GENERAL INFORMATION:
APPLICANT: Schatz, Peter J.
TITLE OF INVENTION: Biocinylation of Proteins
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Kourile and Crew
STREET: One Market Plaza, Stewart Tower
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentia Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/512,983
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/383,753
FILING DATE: 03-FEB-1995
APPLICATION NUMBER: US 08/099,991
FILING DATE: 30-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Smith, William M.
REGISTRATION NUMBER: 30,223
REFERENCE/DOCKET NUMBER: 1038.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-512-983-31

Query Match 75.8% Score 25; DB 4; Length 14;
Best Local Similarity 80.0% Pred. No. 36;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 KMXWR 6
11 11
DB 9 KMLMR 13

Search completed: February 27, 2002, 11:36:09
Job time: 148 sec

82

9

1

.....

.....

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:41:23 ; Search time 66.26 Seconds
(without alignments)
6.708 Million cell updates/sec

Title: US-09-446-109A-21

Perfect score: .33
Sequence: 1 FKKXMR 6

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_1101.*
1: /SIDSB/gcgdata/geneseq/geneseqp/AA1980.DAT.*
2: /SIDSB/gcgdata/geneseq/geneseqp/AA1981.DAT.*
3: /SIDSB/gcgdata/geneseq/geneseqp/AA1982.DAT.*
4: /SIDSB/gcgdata/geneseq/geneseqp/AA1983.DAT.*
5: /SIDSB/gcgdata/geneseq/geneseqp/AA1984.DAT.*
6: /SIDSB/gcgdata/geneseq/geneseqp/AA1985.DAT.*
7: /SIDSB/gcgdata/geneseq/geneseqp/AA1986.DAT.*
8: /SIDSB/gcgdata/geneseq/geneseqp/AA1987.DAT.*
9: /SIDSB/gcgdata/geneseq/geneseqp/AA1988.DAT.*
10: /SIDSB/gcgdata/geneseq/geneseqp/AA1989.DAT.*
11: /SIDSB/gcgdata/geneseq/geneseqp/AA1990.DAT.*
12: /SIDSB/gcgdata/geneseq/geneseqp/AA1991.DAT.*
13: /SIDSB/gcgdata/geneseq/geneseqp/AA1992.DAT.*
14: /SIDSB/gcgdata/geneseq/geneseqp/AA1993.DAT.*
15: /SIDSB/gcgdata/geneseq/geneseqp/AA1994.DAT.*
16: /SIDSB/gcgdata/geneseq/geneseqp/AA1995.DAT.*
17: /SIDSB/gcgdata/geneseq/geneseqp/AA1996.DAT.*
18: /SIDSB/gcgdata/geneseq/geneseqp/AA1997.DAT.*
19: /SIDSB/gcgdata/geneseq/geneseqp/AA1998.DAT.*
20: /SIDSB/gcgdata/geneseq/geneseqp/AA1999.DAT.*
21: /SIDSB/gcgdata/geneseq/geneseqp/AA2000.DAT.*
22: /SIDSB/gcgdata/geneseq/geneseqp/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	32	97.0	30	AAW10360 Hagfish intestinal
2	32	97.0	35	AAW58719 Breast and ovarian
3	30	90.9	63	AAW38597 Human colorectal C
4	29	87.9	21	AAW88765 Human interleukin-
5	29	87.9	21	AAW88766 Human interleukin-
6	29	87.9	21	AAW88767 Human interleukin-
7	29	87.9	29	AAW34430 Gene 44 human secr
8	29	87.9	386	AAW90526 B cell stimulating
9	29	87.9	415	AAW41325 Arabidopsis thaliana
10	29	87.9	431	AAW98178 Arabidopsis thaliana
11	29	87.9	431	AAW41324 Arabidopsis thaliana

12	29	87.9	433	17	AAW04268
13	29	87.9	433	20	AAW42649
14	29	87.9	463	21	AAW41323
15	29	87.9	468	10	AAW90525
16	29	87.9	468	10	AAW90284
17	29	87.9	468	14	AAW37215
18	29	87.9	468	17	AAW98364
19	29	87.9	468	19	AAW71371
20	29	87.9	468	22	AAW36655
21	29	87.9	1068	21	AAW12486
22	29	84.8	61	19	AAW67470
23	28	84.8	324	22	AAW80957
24	28	84.8	330	22	AAW04376
25	28	84.8	330	22	AAW62852
26	28	84.8	330	22	AAW64301
27	28	84.8	330	22	AAW72910
28	28	84.8	330	22	AAW67654
29	28	84.8	495	22	AAW39248
30	28	84.8	498	22	AAW41034
31	28	84.8	597	21	AAW37568
32	28	84.8	799	19	AAW23938
33	28	84.8	1099	19	AAW29683
34	27	81.8	41	22	AAW06638
35	27	81.8	45	22	AAW72953
36	27	81.8	58	22	AAW38400
37	27	81.8	68	21	AAW83243
38	27	81.8	76	21	AAW83237
39	27	81.8	84	21	AAW83237
40	27	81.8	118	20	AAW21845
41	27	81.8	132	21	AAW54631
42	27	81.8	143	21	AAW08276
43	27	81.8	153	21	AAW15526
44	27	81.8	159	22	AAW71907
45	27	81.8	172	21	AAW41238

ALIGNMENTS

RESULT 1	
AAW10360	
ID	AAW10360 standard; peptide: 30 AA.
XX	
NC	AAW10360:
XX	
DT	01-SEP-1997 (first entry)
XX	
DE	Hagfish intestinal antimicrobial peptide 3.
XX	
KW	antibiotic; linear; antimicrobial; hagfish; Myxine glutinosa; Intestine;
KW	bacteria; gram-positive; gram-negative; Fungl; protozoa.
XX	
DS	Myxine glutinosa.
XX	
FIH	
FT	Key
FT	Modified-site 2 Location/Qualifiers
FT	Modified-site 7 /note="Bromine-Trp"
FT	Modified-site 7 /note="Bromine-Trp"
XX	
PN	W09700269-A1.
XX	
PD	03-JAN-1997.
XX	
PE	19-JUN-1996; 96WO-US10801.
XX	
PR	18-JUN-1996; 96US-066543.
XX	
PA	19-JUN-1995; 95US-0000325.
XX	
XX	(MAGA-) MAGANIN PHARM INC.
XX	
XX	Shinnar AE, Uzzeil T, Zasloff MA.
XX	

Ollseed rape seed
Brassica napus Sac
Arabidopsis thaliana
B cell stimulating
Sequence of a rece
IL-6 receptor. Sy
Interleukin-6 rece
Human interleukin-
Human IL-6 recepto
Rice protein for r
Naja nigricollis s
Human nCPCH53. Ho
Human G-protein co
Amino acid sequenc
Human GTP-binding
Human IG53 G-prote
Amino acid sequenc
Human polypeptide
Human polypeptide
Human starch synth
Wheat granule-thion
Human Na-K-2Cl cot
Human foetal prote
Mouse type II inte
Peptide #12437 enc
CYB5Bp fatty acid
CYB5Bp fatty acid
Zea mays protein f
Zea mays signal pept
Zea mays protein f
Amino acid sequenc
Arabidopsis thaliana
Human FV C2 domain
Human ORFX ORF1002

DR WPI: 1997-077486/07.
XX Antimicrobial peptide from hagfish intestine - active against
PT gram-positive and gram-negative bacteria, fungi and protozoa
XX
PS Claim 1; Page 5; 9pp: English.
XX
CC AAM10359-60 are a family of linear antimicrobial peptides from hagfish
CC (Myxine glutinosa) intestine. The peptides have an antimicrobial
CC activity which gives them the ability to kill or disable microorganisms,
CC e.g. bacteria (gram-positive and gram-negative), fungi and protozoa.
XX
XX Sequence: 30 AA:
SO

Query Match 97.0%; Score 32; DB 18; Length 30;
Best Local Similarity 83.3%; Pred. No. 8.7;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 FKKXMR 6
DB 3 FKKXMR 8

RESULT 2
AAB58719
ID AAB58719 standard; Protein: 55 AA.
XX
AC AAB58719;
XX
DT 27-MAR-2001 (first entry)
XX
DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 427.
XX
KW Human: breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW antidiabetic; antiinflammatory; antitumor; vulnerrary; anticonvulsant;
KW antibacterial; antifungal; antiparasitic; cardiac; immune disorder;
KW Addison's disease; allergy; autoimmune haemolytic anaemia;
KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW cardiovascular disorder; wound healing; neurological disease.
XX
OS Homo sapiens.
XX
PN WO200055173-A1.
XX
PD 21-SEP-2000.
XX
PE 08-MAR-2000; 2000MO-US05881.
XX
PK 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI: 2000-611515/58.
DR N-PSDB: AAF21622.
XX
XX
PT New human breast and ovarian cancer associated gene sequences and the
PT polypeptides encoded by these genes, useful in the prevention,
PT treatment and diagnosis of cancer, immune disorders, cardiovascular
PT disorders and neurological diseases -
XX
PS Claim 11; Page 853; 1299pp: English.
XX
XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
CC associated with breast and ovarian cancer. Included in the invention are
CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
CC isolation and characterisation of the DNA and protein sequences of the
CC invention. The breast and ovarian cancer associated DNA, protein, agonist

CC or antagonist sequences exhibit cytostatic; immunosuppressive;
CC neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
CC antidiabetic; antiinflammatory; antitumor; vulnerrary; anticonvulsant;
CC antibacterial; antifungal; antiparasitic and cardiac activity. The
CC polynucleotide and protein sequences are used in the diagnosis of cancer,
CC particularly breast and ovarian cancer. The nucleic acid sequences,
CC prevention and treatment of immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; cardiovascular disorders such as
CC myocardial ischaemias; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases.
XX
XX Sequence 55 AA:
SO

Query Match 97.0%; Score 32; DB 21; Length 55;
Best Local Similarity 83.3%; Pred. No. 16;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 FKKXMR 6
DB 39 FKKXMR 44

RESULT 3
AAM38597
ID AAM38597 standard; Protein: 63 AA.
XX
AC AAM38597;
XX
DT 19-OCT-2001 (first entry)
XX
DE Human colorectal cancer antigen SEQ ID NO: 112.
XX
KW Human: colorectal cancer; colorectal cancer antigen; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200155350-A1.
XX
PD 02-AUG-2001.
XX
PE 17-JAN-2001; 2001WO-US01350.
XX
PK 31-JAN-2000; 2000US-0179065.
PK 04-FEB-2000; 2000US-0180628.
PK 24-FEB-2000; 2000US-0184664.
PK 02-MAR-2000; 2000US-0186350.
PK 16-MAR-2000; 2000US-0189874.
PK 17-MAR-2000; 2000US-0190076.
PK 18-APR-2000; 2000US-0198123.
PK 19-MAY-2000; 2000US-0205515.
PK 07-JUN-2000; 2000US-0209467.
PK 28-JUN-2000; 2000US-0214886.
PK 30-JUN-2000; 2000US-0215135.
PK 07-JUL-2000; 2000US-0216880.
PK 07-JUL-2000; 2000US-0216880.
PK 11-JUL-2000; 2000US-0217487.
PK 11-JUL-2000; 2000US-0217496.
PK 14-JUL-2000; 2000US-0218290.
PK 26-JUL-2000; 2000US-0220963.
PK 26-JUL-2000; 2000US-0220964.
PK 14-AUG-2000; 2000US-0224518.
PK 14-AUG-2000; 2000US-0224519.
PK 14-AUG-2000; 2000US-0225213.
PK 14-AUG-2000; 2000US-0225214.
PK 14-AUG-2000; 2000US-0225266.
PK 14-AUG-2000; 2000US-0225267.
PK 14-AUG-2000; 2000US-0225268.
PK 14-AUG-2000; 2000US-0225270.
PK 14-AUG-2000; 2000US-0225447.

[illegible]

Sequence 65AA:

ID AAB8765 standard; Peptide: 21 AA.
XX
AC AAB8765:
XX
DT 23-MAY-2001. (first entry)
XX
DE Human Interleukin-6 domain IV fragment #6.
XX
KW Interleukin-6; ligand; IL-6 receptor; antagonist; cancer; allergy;
KW rheumatoid arthritis; diabetes; multiple sclerosis; infection;
KW autoimmune disease; inflammatory disease.
XX
OS Homo sapiens.
XX
PN WO200116166-A2.
XX
PD 08-MAR-2001.
XX
PF 25-AUG-2000; 2000WO-US23490.
XX
PR 27-AUG-1999; 99US-0151277.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Saxinger C;
XX
DR WPI: 2001-244395/25.
XX
PT Polypeptides which inhibit the binding of interleukin (IL)-6 ligand
PT with the IL-6 receptor, and the nucleic acids that encode them, useful
PT for treating e.g. inflammation and autoimmune diseases -
XX
PS Example 4; Page 48; 98pp; English.
XX
CC The present invention describes a number of peptides which are able to
CC bind to the human Interleukin-6 (IL-6) receptor and inhibit IL-6 binding.
CC These are useful in the treatment of diseases associated with abnormal
CC IL-6 expression, including multiple myeloma, plasmacytoma, haematological
CC diseases such as plasma cell dyscrasias, leukaemia and lymphoma,
CC mesangial proliferative glomerulonephritis, polyclonal B cell activation
CC conditions, allergies, rheumatoid arthritis, diabetes, multiple
CC sclerosis, septic shock, infections, post-menopausal osteoporosis,
CC chronic immune deficiency, autoimmune diseases and inflammatory diseases.
XX
SQ Sequence 21 AA:

Query Match 87.9%; Score 29; DB 22; Length 21;
Best Local Similarity 66.7%; Pred. No. 23;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKKXMR 6
 | | | |
 13 fkkxw 18
DB

XX
PD 08-MAR-2001.
XX
PF 25-AUG-2000; 2000WO-US23490.
XX
PR 27-AUG-1999; 99US-0151277.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Saxinger C;
XX
DR WPI: 2001-244395/25.
XX
PT Polypeptides which inhibit the binding of interleukin (IL)-6 ligand
PT with the IL-6 receptor, and the nucleic acids that encode them, useful
PT for treating e.g. inflammation and autoimmune diseases -
XX
PS Example 4; Page 48; 98pp; English.
XX
CC The present invention describes a number of peptides which are able to
CC bind to the human Interleukin-6 (IL-6) receptor and inhibit IL-6 binding.
CC These are useful in the treatment of diseases associated with abnormal
CC IL-6 expression, including multiple myeloma, plasmacytoma, haematological
CC diseases such as plasma cell dyscrasias, leukaemia and lymphoma,
CC mesangial proliferative glomerulonephritis, polyclonal B cell activation
CC conditions, allergies, rheumatoid arthritis, diabetes, multiple
CC sclerosis, septic shock, infections, post-menopausal osteoporosis,
CC chronic immune deficiency, autoimmune diseases and inflammatory diseases.
XX
SQ Sequence 21 AA:

Query Match 87.9%; Score 29; DB 22; Length 21;
Best Local Similarity 66.7%; Pred. No. 23;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKKXMR 6
 | | | |
 8 fkkxw 13
DB

RESULT 6
AAB8767
ID AAB8767 standard; Peptide: 21 AA.
XX
AC AAB8767:
XX
DT 23-MAY-2001 (first entry)
XX
DE Human Interleukin-6 domain IV fragment #8.
XX
KW Interleukin-6; ligand; IL-6 receptor; antagonist; cancer; allergy;
KW rheumatoid arthritis; diabetes; multiple sclerosis; infection;
KW autoimmune disease; inflammatory disease.
XX
OS Homo sapiens.
XX
PN WO200116166-A2.
XX
PD 08-MAR-2001.
XX
PF 25-AUG-2000; 2000WO-US23490.
XX
PR 27-AUG-1999; 99US-0151277.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Saxinger C;
XX
DR WPI: 2001-244395/25.
XX
PT Polypeptides which inhibit the binding of interleukin (IL)-6 ligand
PT with the IL-6 receptor, and the nucleic acids that encode them, useful

for treating e.g. inflammation and autoimmune diseases -
 Example 4; Page 48; 98pp; English.

The present invention describes a number of peptides which are able to bind to the human interleukin-6 (IL-6) receptor and inhibit IL-6 binding. These are useful in the treatment of diseases associated with abnormal IL-6 expression, including multiple myeloma, plasmacytoma, haematological diseases such as plasma cell dyscrasias, leukemia and lymphoma, myeloid proliferative glomerulonephritis, polyclonal B cell activation conditions, allergies, rheumatoid arthritis, diabetes, multiple sclerosis, septic shock, infections, post-menopausal osteoporosis, chronic immune deficiency, autoimmune diseases and inflammatory diseases.

Query Match 87.9%; Score 29; DB 22; Length 21;
 Best Local Similarity 66.7%; Pred. No. 23;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FKKXMR 6
 Db 3 FKKLWK 8

RESULT 7

AAB34430
 ID AAB34430 standard; Protein: 29 AA.

AC AAB34430;

DT 26-JAN-2001 (first entry)

DE Gene 44 human secreted protein homologous amino acid sequence #191.

Human; secreted protein; diagnosis; neuroprotective; cytoskeletal;
 KW cardiocastive; immunomodulatory; muscular active general; vulnery;
 KW gastrointestinal; nephrotropic; antineflective; gynaecological; and
 KW antibacterial; gene therapy; detection; cancer; chromosome marker;
 KW chromosome identification; neural disorder; immune disorder;
 KW muscular disorder; reproductive disorder; gastrointestinal disorder;
 KW pulmonary disorder; cardiovascular disorder; renal disorder;
 KW proliferative disorder; wound healing; infectious disease; preservative;
 KW food additive.

OS Acanthophis antarcticus.

PN WO200056883-A1.

PD 28-SEP-2000.

PF 16-MAR-2000; 2000WO-US06822.

PR 23-MAR-1999; 99US-0126054.

PR 10-DEC-1999; 99US-0169916.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM, Komatsoulis G;

DR WPI: 2000-587666/55.

Human secreted proteins and gene sequences encoding them, useful for detecting, preventing, and treating disorders such as cancer, neurological disorders and immune system disorders -
 Disclosure: Page 421; 429pp; English.

The polynucleotide sequences given in AAC59566 to AAC59614 encode the human secreted proteins given in AAB34299 to AAB34347. AAB34348 to AAB34437 represent human secreted polypeptide sequences and proteins homologous to them, which are given in the exemplification of the present

invention. Human secreted proteins have activities based on the tissues and cells the genes are expressed in. Example of activities include: neuroprotective; cytoskeletal; cardiocastive; immunomodulatory; muscular active general; vulnery; gastrointestinal; nephrotropic; antineflective; gynaecological; and antibacterial. The polynucleotides can be used for the detection of various disorders such as cancer, chromosome identification, as chromosome markers, and for numerous other diagnostic or research purposes. The secreted proteins can be used to treat disorders such as neural, immune, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal, and proliferative disorders, wound healing, and infectious diseases. The proteins can also be used as a food additive or preservative to increase or decrease storage capabilities. AAC59557 to AAC59565 and AAB34298 represent sequences used in the exemplification of the present invention.

Query Match 87.9%; Score 29; DB 21; Length 29;
 Best Local Similarity 66.7%; Pred. No. 32;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FKKXMR 6
 Db 14 YKKLWR 19

RESULT 8

AAP90526
 ID AAP90526 standard; Protein: 386 AA.

AC AAP90526;

DT 25-JAN-1990 (first entry)

DE B cell stimulating factor-2 receptor.

B cell stimulating factor-2 receptor; monocytic U937 cell line.

OS Homo sapiens.

PN A08928720-A.

PD 27-JUL-1989.

PF 23-JAN-1989; 89AU-0028720.

PR 22-JAN-1988; 88JP-0012387.

PR 25-JAN-1988; 88JP-0012599.

PR 04-AUG-1988; 88JP-0194885.

TR 14-JAN-1989; 89JP-0007461.

XX (KISH) TADAMITSU KISHIMOTO.

PI Kishimoto T;

DR WPI: 1989-264012/37.

DR N-PSDB: AAP90525.

Receptor protein for human B cell stimulating factor-2 - used for developing prophylactic, therapeutic and diagnostic agents for associated disorders.
 Claim 4; page 37-8; 76pp; english.
 The BSF2 receptor has residues near the N-terminal deleted. The receptor is derived from a monocytic U937 cell line. It can be used to develop prophylactic and therapeutic pharmaceuticals as agents to relate diseases and disorders to abnormal BSF-2 prodn. It can also be used to study an immune mechanism with which BSF-2 or the receptor is concerned.

Sequence 386 AA;

Query Match 87.9% Score 29; DB 10; length 386;
Best Local Similarity 66.7% Pred. No. 4e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 1 FKXMR 6
Db 306 fktckk 311

RESULT 9
AAC41325
ID AAC41325:standard; Protein: 415 AA.
XX
AC AAC41325;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 51401.
XX
KW Protein identification: signal transduction pathway; metabolic pathway;
KM hydrolisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 03-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137328.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.


```

PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158233.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159328.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160960.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match      87.9%: Score 29; DB 21; Length 415;
Best Local Similarity 66.7%: Pred. No. 4.2e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Caps 0;
OY 1 FKXWR 6
III I:

```

```

tbb 71 fKkawk 76
RESULT 10
AAW98178
AAW98178 standard; Protein: 431 AA.
AAW98178:
05-JUL-1999 (first entry)
Anther-specific ESJ2A protein.
ESJ2A gene: promoter; pollen; anther dehiscence; male sterile;
transgenic plant.
Arabidopsis thaliana.
MO9913089-A1.
18-MAR-1999.
11-SEP-1998; 98WO-GB02752.
11-SEP-1997; 97GB-0019359.
(BIOG-) BIOGEMMA UK LTD.
Craze M, Paul W., Roberts JA:
WPI: 1999-254279/21.
N-PSDB; AAX25011.
Generation of male sterile plants by controlling anther dehiscence
Disclosure: Fig 3; 34pp: English.
This protein is encoded by the ESJ2A gene (see AAX25011) of
Arabidopsis thaliana. The invention relates to the use of the
ESJ2A promoter to reduce dehiscence and to create male sterile
plants for use in hybrid seed production. The promoter is used
to drive expression of a further nucleic acid sequence that
results in prevention or reduction of anther dehiscence. For
example, expression of the RNase barnase causes cell ablation,
while expression of a plant hormone alters the developmental fate
of a cell. Plants are produced that have phenotypically normal
pollen grains, within phenotypically normal anthers, but in which
the anthers do not dehiscence and thus do not release the pollen
grains. The system allows the female to be multiplied with the
artificial male sterility gene in the homozygous state, since the
female plant produces viable pollen. The system is suited to
crops which have high seed multiplication. Large amounts of
pollen, and/or separate male and female inflorescences. These
factors allow for easy collection of pollen from non-dehiscing
anthers, lucille self-pollination and the minimisation of the area
of plants that have to be self-pollinated manually. Such an
ideal crop is the monocot maize, but the system is also applicable
to e.g. wheat, barley, rice, fodder grass, banana, palm, orchid,
tulip, lily, melon, cucumber, tomato, pepper and willow. It is
also useful in the avoidance or reduction of pollen allergens and
may be effective in the control of asthma caused by pollen release.
Sequence 431 AA:

```

```

Query Match      87.9%: Score 29; DB 20; Length 431;
Best Local Similarity 66.7%: Pred. No. 4.4e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Caps 0;
OY 1 FKXWR 6
III I:
Db 87 fKkawk 92

```

RESULT 11
AAC41324
ID AAC41324 standard; Protein: 431 AA.
XX AAC41324;
XX
DT 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 51400.
DE
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.

```

PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 29-SEP-1999; 99US-0157117.
PR 04-OCT-1999; 99US-0157753.
PR 05-OCT-1999; 99US-0157865.
PR 06-OCT-1999; 99US-0158029.
PR 07-OCT-1999; 99US-0158232.
PR 08-OCT-1999; 99US-0158369.
PR 12-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match 87.9%; Score 29; DB 21; Length 431;
Best Local Similarity 66.7%; Pred. No. 4.4e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

Oy 1 FKKXWR 6
Db 87 fkkawk 92

```

```

RESULT 12
AAW04268
ID AAW04268 standard; Protein: 433 AA.
XX
AC AAW04268;
XX

```

```

DT 14-DEC-1996 (first entry)
XX
DE Oilseed rape seed pod polygalacturonase.
XX
KW Polygalacturonase; seed pod; dehiscence; shatter; oilseed rape;
XX transgenic plant; antisense.
XX
OS Brassica napus cv. Rafal.
XX
IN WO9630529-A1.
XX
PD 03-OCT-1996.
XX
PF 29-MAR-1996; 96WO-CB00757.
XX
PR 31-MAR-1995; 95GB-0006684.
XX
PA (NICK-) NICKERSON BIOCHEM LTD.
XX
PI Coupe SA, Jenkins ES, Roberts JA;
XX
NR WPI: 1996-455374/45.
XX
NR N-PSDB; AAT31994.
XX
IT Control of seed pod dehiscence - using polygalacturonase or nucleic
XX acid sequences derived from polygalacturonase gene
XX
PS Example 1; Fig 1; 36pp; English.
XX
CC A polygalacturonase (PG) (AAW04248) of oilseed rape cv. Rafal seed
XX pods is useful for controlling dehiscence. Manipulation of the
XX enzyme's activity can influence the timing of dehiscence. A cDNA
XX clone (AAT33994) coding for the PG was isolated from a cDNA library
XX of the rape seed pod dehiscence zone. PG nucleic acids (partic.
XX antisense) can be used to regulate dehiscence in crop plants.
XX
SQ Sequence 433 AA:

```

```

Query Match 87.9%; Score 29; DB 17; Length 433;
Best Local Similarity 66.7%; Pred. No. 4.4e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

Oy 1 FKKXWR 6
Db 88 fkkawk 93

```

```

RESULT 13
AAV42649
ID AAV42649 standard; Protein: 433 AA.
XX
AC AAV42649;
XX
DT 10-JAN-2000 (first entry)
XX
UE Brassica napus Sac66 protein putative sequence.
XX
KW Signal transduction protein; dehiscence; male sterile plant; DZ2 gene;
XX shatter resistance; oilseed rape; Sac66 protein.
XX
US Brassica napus.
XX
PN WO9949046-A1.
XX
PD 30-SEP-1999.
XX
PF 22-MAR-1999; 99WO-CB00905.
XX
PR 20-MAR-1998; 98GB-0006113.
XX
PA (BIOG-) BIOGENMA UK LTD.
XX

```

PI Wyatt P, Roberts JA, Whitelaw C;
XX WPI: 1999-580449/49.
DR N-PSDB: AA222980.
XX A nucleic acid encoding a signal transduction protein involved in plant
PT dehiscence, useful for producing shatter resistant male sterile plants
PT -
XX
PS Example 6: Fig 15; 71pp: English.
XX
CC The invention provides a nucleic acid encoding a signal transduction
CC protein involved in the process of dehiscence. The nucleic acids and
CC proteins are useful for regulating or controlling dehiscence of a pod or
CC an anther in a plant, useful in the production of male sterile plants.
CC The methods, etc., may be used in production of shatter resistance or
CC shatter-delayed plants such as oilseed rape (*Brassica napus*). The present
CC sequence represents a *B. napus* Sac66 protein putative sequence.
XX
SO Sequence 433 AA:

Query Match 87.9%; Score 29; DB 20; Length 433;
Best Local Similarity 66.7%; Pred. No. 4.4e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKKXR 6
DB 88 fkkaw 93

RESULT 14
AAC41323
ID AAC41323 standard; Protein: 463 AA.
XX
AC AAC41323:
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 51399.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 26-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145916.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147403.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.

FR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
FR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161982.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 87.9% Score 29; DB 21; Length 463;
Best Local Similarity 66.7% Pred. NO. 4.7e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Caps 0;

QY 1 FKKXWR 6
DB 119 FKKXWK 124

RESULT 15
AAP90525 standard; protein; 468 AA.

AC AAP90525;

XX 23-JAN-1990 (first entry)

XX B cell stimulating factor-2 receptor.

XX B cell stimulating factor-2 receptor; monocytic U937 cell line.

XX Homo sapiens.

XX AU8928720-A.

XX 27-JUL-1989.

XX 23-JAN-1989; 89AU-0028720.

XX 22-JAN-1988; 88JP-0012387.

XX 25-JAN-1988; 88JP-0012599.

XX 04-AUG-1988; 88JP-0194885.

XX 14-JAN-1989; 89JP-0007461.

XX (KISH) TADAMITSU KISHIMOTO.

XX Kishimoto T;

XX WPI; 1989-264012/37.

XX N-PSDB; AAP90525.

XX Receptor protein for human B cell stimulating factor-2 - used for

XX developing prophylactic, therapeutic and diagnostic agents for

XX associated disorders.

XX Claim 2; page 36; 76pp; english.

XX The BSF2 receptor is derived from a monocytic U937 cell line. It can be

XX used to develop prophylactic and therapeutic pharmaceuticals, as agents to

XX relate diseases and disorders to abnormal BSF-2 prodn. It can also be

XX used to study an immune mechanism with which BSF-2 or the receptor is

XX concerned.

Query Match 87.9% Score 29; DB 10; Length 468;
Best Local Similarity 66.7% Pred. NO. 4.8e+02;

	Matches	4;	Conservative	1;	Mismatches	1;	Indels	0;	Gaps	0;
Qy	1	FKKXWR	6							
	111	1:								
Db	388	fkktwk	393							

Search completed: February 27, 2002, 11:51:50
Job time: 627 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:45:28 ; Search time 34.86 Seconds

(without alignments)
13.111 Million cell updates/sec

Title: US-09-446-109a-21

Perfect score: 33
Sequence: 1 FKKXMR 6Scoring table: BLOSUM62
Gapop 10.0 , Capext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 68: *
1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	97.0	174	2	S43889
2	32	97.0	391	2	S49300
3	32	97.0	397	2	S49301
4	32	97.0	500	2	S49302
5	31	93.9	209	1	C69796
6	31	93.9	348	2	T29663
7	31	93.9	377	2	E75388
8	31	93.9	1191	2	A53491
9	31	93.9	2207	2	T24629
10	29	87.9	60	1	NIEYIH
11	29	87.9	60	2	S68769
12	29	87.9	62	1	N1AW1
13	29	87.9	252	2	H82574
14	29	87.9	431	2	T46187
15	29	87.9	468	1	A41242
16	29	87.9	577	2	E71364
17	29	87.9	726	2	C82548
18	29	87.9	726	2	H82774
19	29	87.9	788	2	T30510
20	29	87.9	3770	2	A40889
21	28	84.8	60	1	NIEPID
22	28	84.8	61	1	NINJ2E
23	28	84.8	61	1	NINJ1C
24	28	84.8	61	1	NINJ1B
25	28	84.8	61	1	NINJ1E
26	28	84.8	61	1	NINJ2C
27	28	84.8	62	1	NINJ1M
28	28	84.8	62	1	NINJ3M
29	28	84.8	62	2	JK0221

30	28	84.8	66	2	B53495	homeobox gene DLX2
31	28	84.8	83	1	NINJ1F	short neurotoxin 1
32	28	84.8	87	2	S08401	neurotoxin homolog
33	28	84.8	122	2	B64493	hypothetical prote
34	28	84.8	198	2	C75005	hypothetical prote
35	28	84.8	236	2	T29953	hypothetical prote
36	28	84.8	262	2	T29954	hypothetical prote
37	28	84.8	270	2	I50111	DLX2 homeodomain p
38	28	84.8	278	2	A56570	homeobox protein D
39	28	84.8	285	2	I51412	hypothetical trans
40	28	84.8	307	2	A45581	distal-less homeob
41	28	84.8	309	2	T29225	hypothetical prote
42	28	84.8	312	2	H83527	hypothetical prote
43	28	84.8	328	2	G02469	homeotic protein D
44	28	84.8	332	2	JH0465	homeotic protein T
45	28	84.8	383	1	C53309	prgy protein - Ent

ALIGNMENTS

RESULT 1

S43889

C:Species: Triticum aestivum (common wheat)

C:Date: 19-Mar-1997 #sequence_revision 01-Aug-1997 #text_change 21-Jul-2000

C:Accession: S43889

R:Jl, L.H.: Langridge, P.

Mol. Gen. Genet. 243, 17-23, 1994

A:Title: An early meiosis cDNA clone from wheat.

A:Reference number: S43889; MUID:94247352

A:Accession: S43889

A:Molecule type: mRNA

A:Residues: 1-174 <JL>

A:Cross-references: EMBL:X79130; NID:9530778; PIDN:CAA55731.1; PID:9530779

A:Experimental source: seed, var. Chinese Spring

C:Superfamily: polygalacturonase-inhibiting protein; leucine-rich alpha-2-glycoprotei

F:9-32/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>

F:127-149/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>

Query Match 97.0% Score 32; DB 2; Length 174;
Best Local Similarity 83.3% Pred. NO. 18;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FKKXMR 6
Db 149 FKKXMR 154

RESULT 2

S49300

C:Species: Triticum aestivum (common wheat)

C:Date: 16-Feb-1995 #sequence_revision 12-May-1995 #text_change 19-May-2000

C:Accession: S49300

R:Langhui, J.; Letarte, J.; Langridge, P.

submitted to the EMBL Data Library, September 1994

A:Reference number: S49299

A:Accession: S49300

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-391 <LIA>

A:Cross-references: EMBL:X81370; NID:9551213; PIDN:CAA57135.1; PID:9551214

C:Superfamily: polygalacturonase-inhibiting protein; leucine-rich alpha-2-glycoprotei

F:114-137/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR1>

F:211-234/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR2>

F:235-258/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>

F:327-350/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>

Query Match 97.0% Score 32; DB 2; Length 391;
Best Local Similarity 83.3% Pred. NO. 39;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 FKKXMR 6
 ||| ||
 Db 350 FKKTMR 355

RESULT 3

AMJ175 protein - wheat
 C:Species: Triticum aestivum (common wheat)
 C:Date: 16-Feb-1995 #sequence_revision 12-May-1995 #text_change 19-May-2000
 C:Accession: S49301
 R:Langhui, J.; Letarte, J.; Langridge, P.
 submitted to the EMBL Data Library, September 1994
 A:Reference number: S49299
 A:Accession: S49301
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-397 <LIn>
 A:Cross-references: EMBL:X81368; NID:q551209; PIDN:CAAS7133.1; PID:q551210
 C:Superfamily: polygalacturanase-inhibiting protein; leucine-rich alpha-2-glycoprotein

Query Match 97.0%; Score 32; DB 2; Length 397;
 Best Local Similarity 83.3%; Pred. No. 39;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FKKXMR 6
 ||| ||
 Db 360 FKKTMR 365

RESULT 4

AMJ1218 protein - wheat
 C:Species: Triticum aestivum (common wheat)
 C:Date: 16-Feb-1995 #sequence_revision 12-May-1995 #text_change 19-May-2000
 C:Accession: S49302
 R:Langhui, J.; Letarte, J.; Langridge, P.
 submitted to the EMBL Data Library, September 1994
 A:Reference number: S49299
 A:Accession: S49302
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-500 <LIn>
 A:Cross-references: EMBL:X81369; NID:q551211; PIDN:CAAS7134.1; PID:q551212
 C:Superfamily: polygalacturanase-inhibiting protein; leucine-rich alpha-2-glycoprotein

Query Match 97.0%; Score 32; DB 2; Length 500;
 Best Local Similarity 83.3%; Pred. No. 48;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FKKXMR 6
 ||| ||
 Db 463 FKKTMR 468

RESULT 5

conserved hypothetical protein yest - Bacillus subtilis
 C:Species: Bacillus subtilis
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
 C:Accession: C69796
 R:Kunst, F.; Ogasawara, N.; Mosser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
 C.; Bron, S.; Broutillet, S.; Bursch, C.V.; Caldwell, B.; Capano, V.; Carter, N.M.; Ch
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Erlingsson, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gall
 lech, J.; Harwood, C.R.; Hensul, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Inilio, M.F.
 Koeller, P.; Kohnstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Laidinois,
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scant
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
 A:Reference number: A69580; MUID:98044033
 A:Accession: C69796
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-209 <KUN>
 A:Cross-references: GB:299107; GB:AL009126; NID:q2632866; PIDN:CAB12513.1; PID:q26330
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: yest
 C:Superfamily: Bacillus subtilis conserved hypothetical protein yest

Query Match 93.9%; Score 31; DB 1; Length 209;
 Best Local Similarity 83.3%; Pred. No. 34;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FKKXMR 6
 ||| ||
 Db 72 FKKEMR 77

RESULT 6

hypothetical protein R08C7.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000
 C:Accession: T29663
 R:Favell, A.; Le, T.T.
 submitted to the EMBL Data Library, July 1996
 A:Description: The sequence of C. elegans cosmid R08C7.
 A:Reference number: 220660
 A:Accession: T29663
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-348 <FAV>
 A:Cross-references: EMBL:U61953; PIDN:AAC48073.1; GSPDB:GM00022; CESP:R08C7.2
 A:Experimental source: strain Bristol N2; clone R08C7
 C:Genetics:
 A:Gene: CESP:R08C7.2
 A:Map position: 4
 A:Insertions: 33/1; 136/3; 167/1; 215/3; 281/1
 C:Superfamily: Saccharomyces cerevisiae probable membrane protein YNR048W

Query Match 93.9%; Score 31; DB 2; Length 348;
 Best Local Similarity 83.3%; Pred. No. 55;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FKKXMR 6
 ||| ||
 Db 250 FKKLMR 255

RESULT 7

conserved hypothetical protein - Deinococcus radiodurans (strain R1)
 C:Species: Deinococcus radiodurans
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
 C:Accession: E75388
 R:White, O.; Eissen, J.A.; Heldelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
 M.; Shen, M.; Vanathavan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896
 A:Accession: E75388

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-377 <WH1>
 A:Cross-references: GB:AE001994; GB:AE000513; NID:96459259; PIDN:AAFI1073.1; PID:9645926
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR1507
 A:Map position: 1

Query Match
 Best Local Similarity 93.9%; Score 31; DB 2; Length 377;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKKXMR 6
 ||| ||
 Db 57 FKKLMR 62

RESULT 8

A53491
 bumetanide-sensitive Na-K-Cl cotransporter - spiny dogfish

C:Species: *Squalus acanthias* (spiny dogfish)
 C>Date: 02-Jun-1994 #sequence_revision 02-Jun-1994 #text_change 02-Mar-2001
 C:Accession: A53491
 R:Xu, J.C.; Lytle, C.; Zhu, T.T.; Payne, J.A.; Benz Jr., E.; Forbush III, B.
 Proc. Natl. Acad. Sci. U.S.A. 91, 2201-2205, 1994
 A:Title: Molecular cloning and functional expression of the bumetanide-sensitive Na-K-Cl
 A:Reference number: A53491; MUID:94181560
 A:Accession: A53491
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1191 <XUA>
 A:Cross-references: GB:U05958; NID:9454096; PIDN:AAB60617.1; PID:9454097
 C:Superfamily: rat bumetanide-sensitive Na+/K+/Cl--cotransport protein

Query Match
 Best Local Similarity 93.9%; Score 31; DB 2; Length 1191;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKKXMR 6
 ||| ||
 Db 861 FKKDMR 866

RESULT 9

T24629
 glutamate synthase (NADH) (EC 1.4.1.14) precursor [similarity] - *Caenorhabditis elegans*

N:Alternate names: protein W07E11.1
 C:Species: *Caenorhabditis elegans*
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
 C:Accession: T24629; T26264
 R:Kershaw, J.
 submitted to the EMBL Data Library, June 1995
 A:Reference number: Z19914
 A:Accession: T24629
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-2207 <WIL>
 A:Cross-references: EMBL:249809; PIDN:CAA90070.1; GSPDB:GN00028; CESP:W07E11.1
 A:Experimental source: clone T06H11
 R:Mortimore, B.
 submitted to the EMBL Data Library, June 1995
 A:Reference number: Z20185
 A:Accession: T26264
 A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA
 A:Residues: 1-2207 <WIL>
 A:Cross-references: EMBL:249868; PIDN:CAA90032.1; GSPDB:GN00028; CESP:W07E11.1
 A:Experimental source: clone W07E11
 C:Genetics:
 A:Gene: CESP:W07E11.1

A:Map position: X
 A:Initons: 19/3; 119/2; 326/3; 389/1; 440/3; 492/3; 583/3; 634/3; 661/1; 801/3; 856/1
 C:Superfamily: glutamate synthase (NADH)
 C:KeyWords: 3fe-4S; metalloprotein; oxidoreductase
 F:1-58/Domain: propeptide #status predicted <PRO>
 F:59-2207/Product: glutamate synthase #status predicted <MAT>
 F:59/Active site: Cys #status predicted
 F:1193,1199,1204/Binding site: 3fe-4S cluster (Cys) (covalent) #status predicted

Query Match
 Best Local Similarity 93.9%; Score 31; DB 2; Length 2207;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKKXMR 6
 ||| ||
 Db 1775 FKKXMR 1780

RESULT 10

NIEX1H
 short neurotoxin 1 - sea snake (*Hydrophis lapemoides*)

C:Species: *Hydrophis lapemoides*
 C>Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 23-Aug-1996
 C:Accession: A90321; A01705
 R:Tamiya, N.; Maeda, N.; Cogger, H.G.
 Biochem. J. 213, 31-38, 1983
 A:Title: Neurotoxins from the venoms of the sea snakes *Hydrophis ornatus* and *Hydrophi*
 A:Reference number: A90321; MUID:83308533
 A:Accession: A90321
 A:Molecule type: protein
 A:Residues: 1-60 <TAM>
 C:Superfamily: snake toxin
 C:KeyWords: postsynaptic neurotoxin; venom
 F:3-22,17-39,41-52,53-58/Disulfide bonds: #status predicted

Query Match
 Best Local Similarity 87.9%; Score 29; DB 1; Length 60;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKKXMR 6
 :|| ||
 Db 23 YKKTMR 29

RESULT 11

S68769
 short neurotoxin - black-banded coral snake

N:Alternate names: alpha-neurotoxin
 C:Species: *Microurus nigrocinctus nigrocinctus* (black-banded coral snake)
 C>Date: 25-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 17-Mar-1999
 C:Accession: S68769
 R:Rosso, J.P.; Vargas-Rosso, O.; Gutierrez, J.M.; Rochat, H.; Bougis, P.E.
 Eur. J. Biochem. 238, 231-239, 1996
 A:Title: Characterization of alpha-neurotoxin and phospholipase A(2) activities from
 in from *Microurus nigrocinctus nigrocinctus*.
 A:Reference number: S68769; MUID:96248443
 A:Accession: S68769
 A:Molecule type: protein
 A:Residues: 1-60 <ROS>
 C:Superfamily: snake toxin
 C:KeyWords: neurotoxin; venom
 F:3-22,17-39,41-52,53-58/Disulfide bonds: #status predicted

Query Match
 Best Local Similarity 87.9%; Score 29; DB 2; Length 60;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKKXMR 6
 :|| ||
 Db 23 YKKTMR 28

RESULT 12

N1AM1

Short neurotoxin 1 - death adder

N:Alternate names: toxin Aa C

C:Species: Acanthophis antarcticus (death adder)

C:Date: 18-Aug-1982 #sequence_revision 18-Aug-1982 #text_change 23-Aug-1996

C:Accession: A01702

R:Kim, H.S.; Tamlye, N.

Biochem. J. 199, 211-218, 1981

A:Title: The amino acid sequence and position of the free thiol group of a short-chain

A:Reference number: A01702; MUID:82160151

A:Accession: A01702

A:Molecule type: protein

A:Residues: 1-62 <IM>

A:Note: Cys-4 is not involved in a disulfide bond

C:Superfamily: snake toxin

C:Keywords: neurotoxin; venom

F:3-24,17-41,43-54,55-60/Disulfide bonds: #status predicted

Query Match 87.9% Score 29; DB 1; Length 62;

Best Local Similarity 66.7% Pred. No. 28;

Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FRKXMR 6

Db 25 YKKTMR 30

RESULT 13

H82574

Phege-related protein XF2291 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C:Accession: H82574

R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: H82574

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-252 <SIM>

A:Cross-references: GB:AE004041; GB:AE003849; NID:99107453; P1DN:AAF85090.1; GSPDB:GN001

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvares, R.; A

Brites, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Carraro, L.E.A.; Carraro, D.M.; Carref, H

as-Melo, E.; Docena, C.; El-Dorri, H.; Facinani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig

Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins, R

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A

Rodrigues, V.; Rosa, A.J. de M.; de Sa, R.G.; Santelli, R.V.; Savasak

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; da Silva, M.A.; da Silveir

M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.;

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF2291

Oy 1 FRKXMR 6

Db 81 FRKXMR 86

RESULT 14

T46187

polygalacturonase (EC 3.2.1.15) precursor [similarity] - Arabidopsis thaliana

N:Alternate names: protein T8H10.110

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 18-Aug-2000

C:Accession: T46187; T50674

R:Benes, V.; Rechmann, S.; Borkova, D.; Ansoorge, W.; Mewes, H.W.; Lemke, K.; Mayer,

submitted to the Protein Sequence Database, January 2000

A:Reference number: 223014

A:Accession: T46187

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-431 <BEN>

A:Cross-references: EMBL:AL133248; P1DN:CA86108.1

A:Experimental source: cultivar Columbia; BAC clone T8H10

R:Jenkins, E.S.; Roberts, J.A.

submitted to the EMBL Data Library, December 1997

A:Description: Dehiscence-related expression of an Arabidopsis thaliana gene encoding

A:Reference number: 225172

A:Accession: T50674

A>Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-431 <JEN>

A:Experimental source: cultivar Landsberg erecta

C:Genetics:

A:Gene: T8H10.110

A:Map position: 3

A:Intons: 85/3; 129/3; 186/3; 193/3; 263/1; 290/2; 326/3; 364/3

C:Superfamily: polygalacturonase

C:Keywords: glycosidase; hydrolase

F:1-27/Domain: signal sequence #status predicted <SIG>

F:28-431/Product: polygalacturonase #status predicted <MAT>

Query Match

Best Local Similarity 87.9% Score 29; DB 2; Length 431;

Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FRKXMR 6

Db 87 FRKXMR 92

RESULT 15

A41242

Interleukin-6 receptor precursor - human

N:Contains: Interleukin-6 receptor, soluble form.

C:Species: Homo sapiens (man)

C:Date: 27-Mar-1992 #sequence_revision 07-Dec-1994 #text_change 21-Jul-2000

C:Accession: A41242; J00080; S17468; A61459; S14621

Yamasaki, K.; Taga, T.; Hirata, Y.; Yawata, H.; Kawanishi, Y.; Seed, B.; Taniguchi,

Science 241, 825-828, 1988

A:Title: Cloning and expression of the human interleukin-6 (BSF-2/IFNbeta 2) receptor

A:Reference number: A41242

A:Accession: A41242

A:Molecule type: mRNA

A:Residues: 1-468 <YAM>

A:Cross-references: GB:M20566; NID:933845; P1DN:CAA31312.1; P1D:933846

Yamasaki, K.; Taga, T.; Hirata, Y.; Yawata, H.; Kawanishi, Y.; Seed, B.; Taniguchi,

Proc. Jpn. Acad. 64, 209-211, 1988

A:Title: Molecular structure of Interleukin 6 receptor.

A:Reference number: J00080

A:Accession: J00080

A:Molecule type: mRNA

A:Residues: 1-468 <YAM>

R:Schmidt, H.; Stoyan, T.; Lenz, D.; Schmitz, H.; Hirano, T.; Kishimoto, T.; Heine

Biochem. J. 277, 659-664, 1991

A:Title: Structural and functional studies on the human hepatic Interleukin-6 recepto

A:Reference number: S17468; MUID:91336983

A:Accession: S17468
A:Molecule type: mRNA
A:Residues: 1-468 <SCH>
A:Cross-references: EMBL:X58298; NID:g32580; PIDN:CAA41231.1; PID:g32581
A:Experimental source: hepatoma cell line HepG2
R:Novick, D.; Engelmann, H.; Wallach, D.; Rubinstein, M.
J. Exp. Med. 170: 1409-1414, 1989
A:Title: Soluble cytokine receptors are present in normal human urine.
A:Reference number: A61459; WUID:90010793
A:Accession: A61459
A:Molecule type: protein
A:Residues: 20-49 <NOV>
C:Comment: Through this receptor, interleukin-6 induces proliferation, activation, and C:Comment: This growth factor receptor does not have a tyrosine kinase domain.
C:Gene: GDB:IL6R
A:Cross-references: GDB:127966; OMIM:147880
A:Map position: 1q21-1q21
C:Superfamily: ciliary neurotrophic factor receptor; cytokine receptor homology; immunog C:Keywords: acute phase; cytokine receptor; glycoprotein; transmembrane protein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-468/Product: interleukin-6 receptor #status predicted <MAT>
F:20-363/Domain: extracellular #status predicted <EXT>
F:40-98/Domain: immunoglobulin homology <IMW2>
F:121-309/Domain: cytokine receptor homology <CRS>
F:364-386/Domain: transmembrane #status predicted <TMW>
F:387-468/Domain: intracellular #status predicted <INT>
F:47-96/Disulfide bonds: #status predicted
F:55,93,221,245,350/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 87.98; Score 29; DB 1; Length 468;
Best Local Similarity 66.7%; Pred. No. 1.8e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 PKKXMR 6
||| |:
Db 388 FKKTWK 393

Search completed: February 27, 2002, 11:52:58
Job time: 450 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:42:53 : Search time 20.39 Seconds
(without alignments)
10.789 Million cell updates/sec

Title: US-09-446-109A-21
Perfect score: 33
Sequence: 1 FKKXMR 6

Scoring table: BLOSUM62
Gapop 10.0 , Capext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Swissprot_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	93.9	741	1	PIFA_ECOLI
2	31	93.9	1191	1	NKCI_SQUAC
3	29	87.9	60	1	NKSI_HYDIA
4	29	87.9	60	1	NKSI_MICNI
5	29	87.9	62	1	NKSI_ACAAN
6	29	87.9	468	1	IL6A_HUMAN
7	29	87.9	577	1	Y110_TREPA
8	29	87.9	3770	1	ACVS_EMENT
9	28	84.8	60	1	NKSI_DENPO
10	28	84.8	61	1	NKSI_NAJHA
11	28	84.8	61	1	NKSI_NAJHA
12	28	84.8	61	1	NKSI_NAJHA
13	28	84.8	61	1	NKSI_NAJHA
14	28	84.8	61	1	NKSI_NAJHA
15	28	84.8	62	1	NKSI_NAJHA
16	28	84.8	62	1	NKSI_NAJHA
17	28	84.8	62	1	NKSI_NAJHA
18	28	84.8	62	1	NKSI_NAJHA
19	28	84.8	66	1	CADO_BUNCA
20	28	84.8	83	1	NKSI_NAJHA
21	28	84.8	87	1	NKSI_NAJHA
22	28	84.8	198	1	YV40_PYRAH
23	28	84.8	270	1	DLX2_BRARE
24	28	84.8	276	1	DLX2_BRARE
25	28	84.8	285	1	HMDA_XENLA
26	28	84.8	328	1	DLX2_HUMAN
27	28	84.8	332	1	DLX2_HUMAN
28	28	84.8	404	1	GLF_MYCE
29	28	84.8	442	1	GNT2_RAT
30	28	84.8	447	1	GNT2_HUMAN
31	28	84.8	495	1	OXAL_HUMAN
32	28	84.8	504	1	XYNB_BACST
33	28	84.8	607	1	VE1_HPV24

34	28	84.8	614	1	VE1_HPV41
35	28	84.8	802	1	YGN9_YEAST
36	28	84.8	1051	1	YC94_HUMAN
37	28	84.8	1095	1	NKCI_MOUSE
38	28	84.8	1095	1	NKCI_MOUSE
39	28	84.8	1099	1	NKCI_MOUSE
40	28	84.8	1099	1	NKCI_MOUSE
41	27	81.8	104	1	Y019_NPOV
42	27	81.8	111	1	Y06R_BPT4
43	27	81.8	202	1	RL13_TOBAC
44	27	81.8	206	1	RL13_TOBAC
45	27	81.8	206	1	RL13_TOBAC

ALIGNMENTS

RESULT 1
ID PIFA_ECOLI STANDARD: PRT: 741 AA.
AC P96329;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PHAGE T7 EXCLUSION PROTEIN.
GN PIFA.
OS Escherichia coli.
OC Plasmid F.
OC Bacteria: Proteobacteria; gamma subdivision: Enterobacteriaceae:
OC Escherichia.
OX NCBI_TaxID=562;
SN (1)
AP SEQUENCE FROM N.A.
RC STRAIN=K12 / CRG3;
FA Shimizu H., Saitoh Y., Suda Y., Uehara K., Sempel G., Mizobuchi K.:
"Complete nucleotide sequence of the F plasmid: its implications for
organization and diversification of plasmid genomes."
Submitted (Apr-2000) to the EMBL/Genbank/DBJ databases.
SN (2)
RP SEQUENCE OF 411-735 FROM N.A.
RC STRAIN=K12;
KA Molineux I.J.:
Submitted (Nov-1996) to the EMBL/Genbank/DBJ databases.
SN (3)
RP FUNCTION.
EX MEDLINE=71240788; PubMed=5283385;
KA Morrison T.G., Malamy M.H.:
"T7 translational control mechanisms and their inhibition by F
factors."
Nature New Biol. 231:37-41(1971).
- FUNCTION: RESPONSIBLE FOR THE EXCLUSION OF PHAGE T7 BY PLASMID F.
GROWTH OF BACTERIOPHAGE T7 IS INHIBITED IN CELLS OF E. COLI THAT
CARRIES THE PLASMID F.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AP001918; BAA97910.1;
CC EMBL: U89941; AAB48033.1;
CC Ecogene: EG40053; PIFA.
CC Plasmid: Complete proteome.
CC CONFLICT 701
CC
CC DOVSPPEDMPAVEELKVLMSGNELTPPEQALKTALEND
CC -> VTRSVRSIMTQOSKRS (IN REF. 2).
CC
CC SEQUENCE 741 AA: 84965 MW: 4C373DF929AFC65 CAC64;
CC
CC Query Match 93.9%; Score 31; DB 1; Length 741;
CC Best Local Similarity 83.3%; Pred. No. 62;


```

RESULT 4
CC NXS1_MICNI STANDARD: PRT: 60 AA.
ID NXS1_MICNI
AC P80548:
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE SHORT NEUROTOXIN 1 (NEUROTOXIN ALPHA).
OS Micurus nigrocinctus (Central American coral snake) (Gargantilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodactylidae; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Elapidae; Elapinae; Micurus.
OX NCBI_TaxID=6635;
RN [1]
RP SEQUENCE.
RC TISSUE-Venom;
RX MEDLINE=96248443; PubMed=8665942;
RA Rosso J.-P., Vargas-Rosso O., Gutierrez J.-M., Rochat H., Bougis P.E.;
RT "Characterization of alpha-neurotoxin and phospholipase A2 activities
RT from Micurus venoms. Determination of the amino acid sequence and
RT receptor-binding ability of the major alpha-neurotoxin from Micurus
RT nigrocinctus nigrocinctus."
RL Eur. J. Biochem. 238:231-239(1996).
CC -1- FUNCTION: PRODUCES PERIPHERAL PARALYSIS BY BLOCKING NEUROMUSCULAR
CC TRANSMISSION AT THE POSTSYNAPTIC SITE. BINDS TO THE NICOTINIC
CC ACETYLCHOLINE RECEPTOR.
DR HSSP: P01426; INEA.
DR InterPro: IPR003571; Snake_toxin.
DR Pfam: PF00087; Toxin_1.
DR ProDom: PD000206; Snake_toxin_1.
DR ProSite: PS00272; SNAKE_TOXIN_1.
KM Venom: Neurotoxin; Postsynaptic neurotoxin.
FT DISULFID 3 22 BY SIMILARITY.
FT DISULFID 17 39 BY SIMILARITY.
FT DISULFID 41 52 BY SIMILARITY.
FT DISULFID 53 58 BY SIMILARITY.
SO SEQUENCE 60 AA; 6582 MW; 4E9580F866F0279C CRC64;

Query Match 87.9%; Score 29; DB 1; Length 60;
Best Local Similarity 66.7%; Pred. No. 15;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

CY 1 FKKXMR 6
DB 23 YKKTWR 28

RESULT 5
CC NXS1_ACAAN STANDARD: PRT: 62 AA.
ID NXS1_ACAAN
AC P01434;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE SHORT NEUROTOXIN 1 (TOXIN AA C).
OS Acanthophis antarcticus (Common death adder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodactylidae; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Elapidae; Acanthophiinae; Acanthophis.
OX NCBI_TaxID=8605;
RN [1]
RP SEQUENCE.
RC TISSUE-Venom;
RX MEDLINE=82160151; PubMed=7337702;
RA Kim H.S., Tamiya N.;
RT "The amino acid sequence and position of the free thiol group of a
RT short-chain neurotoxin from common-death-adder (Acanthophis
RT antarcticus) venom."
RL Biochem. J. 199:211-218(1981).
CC -1- FUNCTION: PRODUCES PERIPHERAL PARALYSIS BY BLOCKING NEUROMUSCULAR
CC TRANSMISSION AT THE POSTSYNAPTIC SITE. BINDS TO THE NICOTINIC

```

```

CC ACETYLCHOLINE RECEPTOR.
CC -1- MISCELLANEOUS: LD(50) IS 0.08 MG/KG BY INTRAMUSCULAR INJECTION.
DR PIR: A01702; NIAMI.
DR HSSP: P01426; INEA.
DR InterPro: IPR003571; Snake_toxin.
DR Pfam: PF00087; Toxin_1.
DR ProDom: PD000206; Snake_toxin_1.
DR ProSite: PS00272; SNAKE_TOXIN_1.
KM Venom: Neurotoxin; Postsynaptic neurotoxin; Multigene family.
FT DISULFID 3 24 BY SIMILARITY.
FT DISULFID 17 41 BY SIMILARITY.
FT DISULFID 43 54 BY SIMILARITY.
FT DISULFID 55 60 BY SIMILARITY.
SO SEQUENCE 62 AA; 6880 MW; 2C1C29E397E84DB2 CRC64;

Query Match 87.9%; Score 29; DB 1; Length 62;
Best Local Similarity 66.7%; Pred. No. 15;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

CY 1 FKKXMR 6
DB 25 YKKTWR 30

RESULT 6
CC IL6A_HUMAN STANDARD: PRT: 468 AA.
ID IL6A_HUMAN
AC P08887; Q16202;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE INTERLEUKIN-6 RECEPTOR ALPHA CHAIN PRECURSOR (IL-6R-ALPHA) (IL-6R 1)
DE (CD126 ANTIGEN).
NM IL6R.
CS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (LONG ISOFORM).
RX MEDLINE=88305347; PubMed=3136546;
RA Yamasaki K., Taga T., Hirata Y., Yawata H., Kawanishi Y., Seed B.,
RA Taniguchi T., Hirano T., Kishimoto T.;
RT "Cloning and expression of the human interleukin-6 (BSF-2/IFN beta 2)
RT receptor."
RT Science 241:825-828(1988).
PN [2]
RP SEQUENCE FROM N.A. (LONG ISOFORM).
KA Yamasaki K., Taga T., Hirata Y., Yawata H., Kawanishi Y., Seed B.,
KA Taniguchi T., Hirano T., Kishimoto T.;
RT "Molecular structure of interleukin 6 receptor."
RT Proc. Jpn. Acad. B, Phys. Biol. Sci. 64:209-211(1988).
RN [3]
RP SEQUENCE FROM N.A. (LONG ISOFORM).
RX MEDLINE=91336983; PubMed=1872801;
KA Schooflink H., Stoyan T., Lenz D., Schmitz H., Hirano T.,
KA Kishimoto T., Heintich P.C., Rose-John S.;
RT "Structural and functional studies on the human hepatic interleukin-6
RT receptor. Molecular cloning and overexpression in HepG2 cells."
RL Biochem. J. 277:659-664(1991).
RN [4]
RP SEQUENCE OF 313-365 FROM N.A. (SHORT ISOFORM).
RX MEDLINE=94333499; PubMed=8056053;
RA Horiuchi S., Koyanagi Y., Zhou Y., Miyamoto H., Tanaka Y., Waki M.,
RA Matsumoto A., Yamamoto M., Yamamoto N.;
RT "Soluble interleukin-6 receptors released from T cell or
RT granulocyte/macrophage cell lines and human peripheral blood
RT mononuclear cells are generated through an alternative splicing
RT mechanism."
RL Eur. J. Immunol. 24:1945-1948(1994).
RN [5]
RP PARTIAL SEQUENCE, CARBOHYDRATE-LINKAGE SITES, AND DISULFIDE BONDS.

```

RX MEDLINE-99167486: PubMed-10066782:
 RA Cole A.R., Hall N.E., Treutlein H.R., Eddes J.S., Reid G.E.,
 RA Moritz R.L., Simpson R.J.:
 RT "Disulfide bond structure and N-glycosylation sites of the
 RT extracellular domain of the human Interleukin-6 receptor."
 RL J. Biol. Chem. 274:7207-7215(1999).
 RN 161
 RP MUTAGENESIS.
 RX MEDLINE-93223711: PubMed-8467812:
 RA Yawata H., Yasukawa K., Natsuka S., Murakami M., Yamasaki K., Hibl M.,
 RA Taga T., Kishimoto T.:
 RT "Structure-function analysis of human IL-6 receptor: dissociation of
 RT amino acid residues required for IL-6-binding and for IL-6 signal
 RT transduction through gp130."
 RL EMBO J. 12:1705-1712(1993).
 CC -1- FUNCTION: PART OF THE RECEPTOR FOR INTERLEUKIN 6. BINDS TO IL-6
 CC WITH LOW AFFINITY, BUT DOES NOT TRANSDUCE A SIGNAL. SIGNAL
 CC ACTIVATION NECESSITATE AN ASSOCIATION WITH IL6ST. ACTIVATION MAY
 CC LEAD TO THE REGULATION OF THE IMMUNE RESPONSE. ACUTE-PHASE
 CC REACTIONS AND HEMATOPOIESIS.
 CC -1- FUNCTION: LOW CONCENTRATION OF A SOLUBLE FORM OF INTERLEUKIN-6
 CC RECEPTOR ACTS AS AN AGONIST OF IL-6 ACTIVITY.
 CC -1- SUBUNIT: HEXAMER OF 2 IL6, 2 IL6R-ALPHA AND 2 IL6ST.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND A
 CC SECRETED (SHORT FORM).
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM: ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: THE SHORT FORM IS EXPRESSED IN PERIPHERAL
 CC BLOOD MONONUCLEAR CELLS AND WEAKLY FOUND IN URINE AND SERUM.
 CC -1- DOMAIN: THE TWO FIBRONECTIN TYPE III-LIKE DOMAINS CONTAINED IN THE
 CC N-TERMINAL PART FORM TOGETHER A CYTOKINE-BINDING DOMAIN.
 CC -1- PFM: A SHORT SOLUBLE FORM MAY ALSO BE RELEASED FROM THE MEMBRANE
 CC BY PROTEOLYSIS.
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.
 CC -1- DATABASE: NAME-PROW; NOTE-CD guide CD126 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd126.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X12830: CAA31312.1: -
 DR EMBL: X58298: CAA41231.1: -
 DR EMBL: S72848: AAC60635.1: -
 DR PIR: A41242: A41242.
 DR PIR: J00080: J00080.
 DR PIR: S14621: S14621.
 DR PIR: S17468: S17468.
 DR MIM: 147880: -
 DR InterPro: IPR002996: CRI1A.
 DR InterPro: IPR001777: FN_III.
 DR InterPro: IPR003530: Hematopo_rceptor_L_F3.
 DR InterPro: IPR003006: Ig_MHC.
 DR InterPro: IPR003598: Ig_C2.
 DR Pfam: PF00041: fn3.1.
 DR Pfam: PF00047: Ig.1.
 DR SMART: SM00060: FN3.1.
 DR SMART: SM00408: IG_C2.1.
 DR PROSITE: PS01354: HEMATOPO_REC_L_F3.1.
 DR Receptor: Transmembrane; Glycoprotein; Immunoglobulin domain; Repeat;
 KW Alternative splicing; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 468 INTERLEUKIN-6 RECEPTOR ALPHA CHAIN.
 FT DOMAIN 20 365 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 365 386 POTENTIAL.
 FT DOMAIN 387 468 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 40 103 IG-LIKE C2-TYPE DOMAIN.

Query Match	Best Local Similarity	Score 29: DB 1: Length 468:
1 EKKXMR 6	66.7%	Pred. No. 99:
368 FKKTWK 393	1: Mismatches	1: Indels 0: Gaps 0:
DB	4: Conservative	
RESULT 7		
Y110_TREPA	STANDARD: PRT: 577 AA.	
ID Y110_TREPA		
AC O83148:		
DT 15-DEC-1998 (Rel. 37, Created)		
DT 15-DEC-1998 (Rel. 37, Last sequence update)		
DT 20-AUG-2001 (Rel. 40, Last annotation update)		
DE HYPOTHETICAL PROTEIN TP0110.		
CN TP0110.		
OS Treponema pallidum.		
OC Bacteria: Spirochaetales: Treponema.		

FT DISULFID 25 193
 FT DISULFID 47 96
 FT DISULFID 121 132
 FT DISULFID 165 176
 FT CARBOHYD 55 55
 FT CARBOHYD 93 93
 FT CARBOHYD 221 221
 FT VARSPLIC 356 365
 FT VARSPLIC 366 468
 FT VARSPLIC 121 121
 FT MUTAGEN 122 122
 FT MUTAGEN 132 132
 FT MUTAGEN 134 134
 FT MUTAGEN 140 140
 FT MUTAGEN 153 153
 FT MUTAGEN 165 165
 FT MUTAGEN 174 174
 FT MUTAGEN 176 176
 FT MUTAGEN 184 184
 FT MUTAGEN 190 190
 FT MUTAGEN 193 193
 FT MUTAGEN 211 211
 FT MUTAGEN 217 217
 FT MUTAGEN 232 232
 FT MUTAGEN 233 233
 FT MUTAGEN 254 254
 FT MUTAGEN 277 277
 FT MUTAGEN 278 278
 FT MUTAGEN 279 279
 FT MUTAGEN 280 280
 FT MUTAGEN 281 281
 FT MUTAGEN 285 285
 FT MUTAGEN 291 291
 FT MUTAGEN 293 293
 FT SEQUENCE 468 AA; 51547 MW; 62AA219FA14F1BB8 CRC64;

N-LINKED (GLCNAC. . .)
 N-LINKED (GLCNAC. . .)
 N-LINKED (GLCNAC. . .)
 VOSSSYPLP -> GSRRRSCL (IN SHORT
 ISOFORM).
 MISSING (IN SHORT ISOFORM).
 C->S: COMPLETE LOSS OF LIGAND-BINDING.
 F->A: NO CHANGE OF LIGAND-BINDING AND IL6
 SIGNALING.
 C->A: COMPLETE LOSS OF LIGAND-BINDING.
 W->L: COMPLETE LOSS OF LIGAND-BINDING.
 W->G: NO CHANGE OF LIGAND-BINDING AND IL6
 SIGNALING.
 F->L: NO CHANGE OF LIGAND-BINDING AND IL6
 SIGNALING.
 C->L: COMPLETE LOSS OF LIGAND-BINDING.
 F->L: NO CHANGE OF LIGAND-BINDING AND IL6
 SIGNALING.
 C->A: COMPLETE LOSS OF LIGAND-BINDING.
 D->T: 30% DECREASE OF LIGAND-BINDING AND
 IL6 SIGNALING.
 V->G: 80% DECREASE OF LIGAND-BINDING AND
 NO IL6 SIGNALING.
 C->D: COMPLETE LOSS OF LIGAND-BINDING.
 C->A: NO CHANGE OF LIGAND-BINDING AND IL6
 SIGNALING.
 D->V: COMPLETE LOSS OF LIGAND-BINDING.
 R->S: 30% DECREASE OF LIGAND-BINDING AND
 IL6 SIGNALING.
 W->Q: 30% DECREASE OF LIGAND-BINDING AND
 INCREASE OF IL6 SIGNALING.
 E->A: 50% DECREASE OF LIGAND-BINDING AND
 IL6 SIGNALING.
 C->D: 30% INCREASE OF LIGAND-BINDING AND
 100% INCREASE IN IL6 SIGNALING.
 V->N: 50% DECREASE OF LIGAND-BINDING AND
 50% INCREASE IN IL6 SIGNALING.
 I->D: COMPLETE LOSS OF LIGAND-BINDING.
 H->I: NO CHANGE OF LIGAND-BINDING AND NO
 IL6 SIGNALING.
 D->G: 70% DECREASE OF LIGAND-BINDING AND
 NO IL6 SIGNALING.
 G->D: 80% DECREASE OF LIGAND-BINDING AND
 NO IL6 SIGNALING.
 O->K: COMPLETE LOSS OF LIGAND-BINDING.
 R->G: COMPLETE LOSS OF LIGAND-BINDING.

OX NCBI_Taxid=160;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NICHOLS;
 RX MEDLINE=98332770; PubMed=9665876;
 RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.C.,
 RA Dodson R., Gwynn M., Hickey E.R., Clayton R., Ketchum K.A.,
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utechtack T.,
 RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujil C., Garland S.,
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
 RA Venter J.C.;
 RT "Complete genome sequence of Treponema pallidum, the syphilis
 RT spirochete."
 RL Science 281:375-388(1998).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE001196; AAC65108.1;
 DR TIGR: TP0110;
 DR KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 577 AA; 65149 MW; 55735A7173C3FF98 CRC64;

Query Match 87.9% Score 29; DB 1; Length 577;
 Best Local Similarity 66.7%; Pred. No. 1.2e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FKKXWR 6
 1111
 Db 345 FKKXWK 350

RESULT 8
 ACVS_Emeni STANDARD; PRT: 3770 AA.
 ID ACVS_Emeni
 AC P27742;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE DELTA-(L-ALPHA-AMINOADIPYL)-L-CYSTEINYL-D-VALINE SYNTHETASE
 DE (EC 6.3.1.1) (ACV SYNTHETASE) (ACVS).
 GN ACVA.
 OS Emericella nidulans (Aspergillus nidulans).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; Emericella.
 OX NCBI_Taxid=5072;
 RN [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RC STRAIN=GI91;
 RX MEDLINE=91286299; PubMed=2061333;
 RA Macabee A.P., van Liepelt H., Pallisa H., Unkles S.E., Riach M.B.R.,
 RA Pfeiffer E., von Doehren H., Klinghorn J.R.;
 RT "Delta-(L-alpha-aminoadipyl)-L-cysteiny-D-valine synthetase from
 RT Aspergillus nidulans. Molecular characterization of the acva gene
 RT encoding the first enzyme of the penicillin biosynthetic pathway.";
 RT J. Biol. Chem. 266:12646-12654(1991).
 CC -i- FUNCTION: EACH OF THE CONSTITUENT AMINO ACIDS OF THE TRIPEPTIDE
 CC ACV ARE ACTIVATED AS AMINOADIPYL-ADENYLATES WITH PEPTIDE BONDS
 CC FORMED THROUGH THE PARTICIPATION OF AMINO ACID THIOLESTER
 CC INTERMEDIATES.
 CC -i- CORCTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTHETINES
 CC (POTENTIAL).
 CC -i- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF PENICILLIN AND
 CC CEPHALOSPORIN.
 CC -i- PTM: THE N-TERMINUS IS BLOCKED.
 CC -i- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME

FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X54853; CAA38631.1;
 DR PIR: A40889;
 DR HSSP: P14687; 1AMU;
 DR InterPro: IPR000873; AMP-bind.
 DR InterPro: IPR001242; DGF4.
 DR InterPro: IPR000379; Est_1lp_thioest_actsite.
 DR InterPro: IPR003880; Phosphopant_attach.
 ER InterPro: IPR001031; Thioesterase.
 ER Pfam: PF00501; AMP-binding_3.
 DR Pfam: PF00668; Condensation; 3.
 DR Pfam: PF00550; PP-binding; 3.
 DR Pfam: PF00975; Thioesterase; 1.
 DR PRINTS: PR00154; AMPBINDING.
 DR PROSITE: PS00012; PHOSPHOPANTHETINE; 3.
 DR PROSITE: PS00455; AMP-BINDING; 3.
 DR PROSITE: PS0075; ACP-MAIN; 3.
 EV Ligase; Antibiotic biosynthesis; Multifunctional enzyme;
 KW Repeat; Phosphopantetheine.
 FT REPEAT 321 910 DOMAIN 1 (ADIPATE-ACTIVATING).
 FT REPEAT 1413 1993 DOMAIN 2 (CYSTEINE-ACTIVATING).
 FT REPEAT 2494 3078 DOMAIN 3 (VALINE-ACTIVATING).
 FT DOMAIN 850 919 ACYL CARRIER (ACP) 1.
 FT DOMAIN 1929 2002 ACYL CARRIER (ACP) 2.
 FT BINDING 3020 3087 ACYL CARRIER (ACP) 3.
 FT BINDING 882 882 PHOSPHOPANTHETINE (BY SIMILARITY).
 FT BINDING 1965 1965 PHOSPHOPANTHETINE (BY SIMILARITY).
 FT BINDING 3050 3050 PHOSPHOPANTHETINE (BY SIMILARITY).
 FT ACT SITE 3623 3623 THIOESTERASE (BY SIMILARITY).
 SQ SEQUENCE 3770 AA; 422448 MW; C86B6D32A58C80 CRC64;

Query Match 87.9% Score 29; DB 1; Length 3770;
 Best Local Similarity 66.7%; Pred. No. 6.9e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FKKXWR 6
 1111
 Db 988 FKKXWK 993
 RESULT 9
 IIXS1_DENPO STANDARD; PRT: 60 AA.
 ID IIXS1_DENPO
 AC P01416;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE SHORT NEUROTOXIN 1 (NEUROTOXIN ALPHA).
 OS Dendroaspis polylepis polylepis (Black mamba).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
 OC Elapidae; Elapinae; Dendroaspis.
 OX NCBI_Taxid=8620;
 RN [1]
 RP TISSUE=Venom;
 RC TISSUE=Venom;
 RX MEDLINE=72206049; PubMed=5033401;
 RA Stridom D.J.;
 RT "Snake venom toxins. The amino acid sequences of two toxins from
 RT Dendroaspis polylepis polylepis (black mamba) venom.";
 RT J. Biol. Chem. 247:4029-4042(1972).
 RN [2]
 RP STRUCTURE BY NMR.

```

RX MEDLINE=89052698; PubMed=2847926;
RA Labhardt A.L., Hunziker-Kwik E.H., Wuehrlich K.;
RT "Secondary structure determination for alpha-neurotoxin from
RT Dendroaspis polylepis polylepis based on sequence-specific
RT 1H-nuclear-magnetic-resonance assignments.";
RL Eur. J. Biochem. 177:295-305(1988).
RN [3]
RP STRUCTURE BY NMR.
RX MEDLINE=93059364; PubMed=1433289;
RA Brown L.R., Wuehrlich K.;
RT "Nuclear magnetic resonance solution structure of the
RT alpha-neurotoxin from the black mamba (Dendroaspis polylepis
RT polylepis).";
RL J. Mol. Biol. 227:1118-1135(1992).
CC -1- FUNCTION: PRODUCES PERIPHERAL PARALYSIS BY BLOCKING NEUROMUSCULAR
CC TRANSMISSION AT THE POSTSYNAPTIC SITE. BINDS TO THE NICOTINIC
CC ACETYLCHOLINE RECEPTOR.
CC -1- MISCELLANEOUS: LD(50) IS 0.09 MG/KG BY SUBCUTANEOUS INJECTION.
DR PIR: A01866; NINEPID.
DR PIR: S01797; S01797.
DR PDB: INTX: 31-JAN-94.
DR InterPro: IPR003571; Snake_toxin.
DR Pfam: PF00087; toxin; 1.
DR ProDom: PD000206; Snake_toxin; 1.
DR PROSITE: PS00272; SNAKE_TOXIN; 1.
KM Venom; Neurotoxin; Postsynaptic neurotoxin; Multigene family;
KW 3D-structure.
FT DISULFID 3 22
FT DISULFID 17 39
FT DISULFID 41 52
FT DISULFID 53 58
FT STRAND 2 4
FT TURN 8 9
FT STRAND 14 16
FT STRAND 22 28
FT STRAND 33 39
FT TURN 46 47
FT STRAND 49 53
FT TURN 56 57
SQ SEQUENCE 60 AA: 6915 MW: E68D6A4410645AC CRC64:

Query Match 84.8%; Score 28; DB 1; Length 60;
Best Local Similarity 66.7%; Pred. No. 23;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FKKXMR 6
DB 23 YKKYWR 28

RESULT 10
NXSL_NAJHA STANDARD: PRT: 61 AA.
AC P01429;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE SHORT NEUROTOXIN 1 (NEUROTOXIN ALPHA) (TOXIN CM-6) (NEUROTOXIN DELTA).
OS Naja naja annulifera (Banded Egyptian cobra),
OS Naja naja naja (Egyptian cobra), and Naja naja (Cape cobra).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Elapidae; Elapinae; Naja.
OX NCBI_TaxID=8641, 8642, 8655;
RN [1]
RP SEQUENCE (ALPHA).
RX SPECIES=N.h.annulifera; TISSUE=Venom;
RC MEDLINE=69257191; PubMed=4308169;
RA Botes D.P., Strydom D.J.;
RP "A neurotoxin, toxin alpha, from Egyptian cobra (Naja naja naja)
RT venom. I. Purification, properties, and complete amino acid
RT sequence.";
```

```

PL J. Biol. Chem. 244:4147-4157(1969).
RN [2]
RN SEQUENCE.
RP SPECIES=N.n. naja; TISSUE=Venom;
RC MEDLINE=79062487; PubMed=718974;
RX Joubert F.J., Taljaard N.;
RT "Purification, some properties and the primary structures of three
RT reduced and S-carboxymethylated toxins (CM-5, CM-6 and CM-10a) from
RT Naja naja naja (Egyptian cobra) venom.";
RL Biochim. Biophys. Acta 537:1-8(1978).
RN [3]
RP SEQUENCE (DELTA).
RC SPECIES=N.nivea; TISSUE=Venom;
RX MEDLINE=71185682; PubMed=4995745;
RA Botes D.P., Strydom D.J., Anderson C.G., Christensen P.A.;
RT "Snake venom toxins. Purification and properties of three toxins from
RT Naja nivea (Linnaeus) (Cape cobra) venom and the amino acid sequence
RT of toxin delta.";
RL J. Biol. Chem. 246:3132-3139(1971).
CC -1- FUNCTION: PRODUCES PERIPHERAL PARALYSIS BY BLOCKING NEUROMUSCULAR
CC TRANSMISSION AT THE POSTSYNAPTIC SITE. BINDS TO THE NICOTINIC
CC ACETYLCHOLINE RECEPTOR.
CC -1- MISCELLANEOUS: LD(50) IS 0.105 MG/KG BY SUBCUTANEOUS INJECTION FOR
CC ALPHA, 0.10 MG/KG BY INTRAVENOUS INJECTION FOR CM-6, AND 0.086
CC MG/KG BY INTRAVENOUS INJECTION FOR DELTA.
DR PIR: A01697; NINJIE.
DR PIR: A92084; NINJ2C.
DR HSSP: P01426; INEA.
DR InterPro: IPR003571; Snake_toxin.
DR Pfam: PF00087; toxin; 1.
DR ProDom: PD000206; Snake_toxin; 1.
DR PROSITE: PS00272; SNAKE_TOXIN; 1.
KM Venom; Neurotoxin; Postsynaptic neurotoxin; Multigene family.
KW BY SIMILARITY.
FT DISULFID 3 23
FT DISULFID 17 40
FT DISULFID 42 53
FT DISULFID 54 59
FT BY SIMILARITY.
SQ SEQUENCE 61 AA: 6844 MW: 7504536B8E96D676 CRC64:

Query Match 84.8%; Score 28; DB 1; Length 61;
Best Local Similarity 66.7%; Pred. No. 23;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FKKXMR 6
DB 24 YKKRWR 29

RESULT 11
NXSL_NAJPA STANDARD: PRT: 61 AA.
AC P01426;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE SHORT NEUROTOXIN 1 (NEUROTOXIN ALPHA).
OS Naja pallida (Red spitting cobra).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Elapidae; Elapinae; Naja.
OX NCBI_TaxID=8658;
RN [1]
RP SEQUENCE.
RC TISSUE=Venom;
RA Baker D.L., Porath J.;
RT "The amino acid sequence of a neurotoxin from Naja nigricollis
RT venom.";
RL Jpn. J. Microbiol. 11:353-355(1967).
RN [2]
RP STRUCTURE BY NMR.
RX MEDLINE=93075732; PubMed=1332755;
RA Zinn-Justin S., Roumestand C., Gillaudin B., Bontems F., Menez A.,
```

RA Tome F.:
 RT "Three-dimensional solution structure of a curarimimetic toxin from
 RL Naja nigricollis venom: a proton NMR and molecular modeling study."
 CC Biochemistry 31:11335-11347(1992).
 CC -1- FUNCTION: PRODUCES PERIPHERAL PARALYSIS BY BLOCKING NEUROMUSCULAR
 CC TRANSMISSION AT THE POSTSYNAPTIC SITE. BINDS TO THE NICOTINIC
 CC ACETYLCHOLINE RECEPTOR.
 CC -1- MISCELLANEOUS: LD(50) IS 0.036 MG/KG BY SUBCUTANEOUS INJECTION.
 CC -1- CAUTION: THE VENOM OF THIS SNAKE WAS ORIGINALLY THOUGHT TO BE THAT
 CC OF N. NIGRICOLLIS.
 CC PIR: A01695; NINJIB.
 DR PDB: 1NEA; 31-OCT-93.
 DR InterPro: IPR003571; Snake_toxin.
 DR Pfam: PF00087; Toxin: 1.
 DR ProDom: PD000206; Snake_toxin: 1.
 DR ProSITE: PS00272; SNAKE_TOXIN: 1.
 KW Venom; Neurotoxin; Postsynaptic neurotoxin; Multigene family;
 KW 3D-structure.
 FT DISULFID 3 23
 FT DISULFID 17 40
 FT DISULFID 42 53
 FT DISULFID 54 59
 FT STRAND 2 4
 FT TURN 8 9
 FT STRAND 14 16
 FT TURN 18 19
 FT STRAND 23 29
 FT STRAND 34 40
 FT TURN 47 48
 FT STRAND 49 54
 SO SEQUENCE 61 AA: 6795 MW: 75186011221C3C1 CRC64;

Query Match 84.8%; Score 28; DB 1; Length 61;
 Best Local Similarity 66.7%; Pred. No. 23;
 Matches 4: Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKKXMR 6
 : 11 11
 Db 24 YKKWVR 29

RESULT 12
 NX52_NAJHA STANDARD: PRT: 61 AA.
 AC P01422:
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-MAY-1992 (Rel. 22, Last annotation update)
 DE SHORT NEUROTOXIN 2 (TOXIN CM-14) (TOXIN V-N-12).
 OS Naja haie annulifera (Banded Egyptian cobra).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
 OC Elapidae; Elapinae; Naja.
 OC NCBI_TaxID=8641;
 OX [1]
 RN SEQUENCE.
 RP TISSUE-Venom;
 RC Joubert F.J.;
 RA "The amino acid sequences of three toxins (CM-10, CM-12 and CM-14)
 RT from Naja haie annulifera (Egyptian cobra) venom."
 RL Hoppe-Seyler's Z. Physiol. Chem. 356:53-72(1975).
 CC -1- FUNCTION: PRODUCES PERIPHERAL PARALYSIS BY BLOCKING NEUROMUSCULAR
 CC TRANSMISSION AT THE POSTSYNAPTIC SITE. BINDS TO THE NICOTINIC
 CC ACETYLCHOLINE RECEPTOR.
 CC -1- MISCELLANEOUS: LD(50) IS 0.12 MG/KG BY SUBCUTANEOUS INJECTION.
 CC PIR: A01691; NINJZE.
 DR HSSP: P01426; 1NEA.
 DR InterPro: IPR003571; Snake_toxin.
 DR Pfam: PF00087; toxin: 1.
 DR ProDom: PD000206; Snake_toxin: 1.
 DR ProSITE: PS00272; SNAKE_TOXIN: 1.
 KW Venom; Neurotoxin; Postsynaptic neurotoxin; Multigene family.

FT DISULFID 3 23 BY SIMILARITY.
 FT DISULFID 17 40 BY SIMILARITY.
 FT DISULFID 42 53 BY SIMILARITY.
 FT DISULFID 54 59 BY SIMILARITY.
 SO SEQUENCE 61 AA: 6915 MW: B2217095FA3C0F0 CRC64;

Query Match 84.8%; Score 28; DB 1; Length 61;
 Best Local Similarity 66.7%; Pred. No. 23;
 Matches 4: Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKKXMR 6
 : 11 11
 Db 24 YKKWVR 29

RESULT 13
 NX52_NAJHH STANDARD: PRT: 61 AA.
 AC P23673:
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE SHORT NEUROTOXIN 2 (TOXIN CM-10A).
 OS Naja haie haie (Egyptian cobra).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
 OC Elapidae; Elapinae; Naja.
 OC NCBI_TaxID=8642;
 OX [1]
 RN SEQUENCE.
 RP TISSUE-Venom;
 RC MEDLINE=79062487; PubMed=718974;
 RX Joubert F.J.; Taljaard N.;
 RT "Purification, some properties and the primary structures of three
 RT reduced and S-carboxymethylated toxins (CM-5, CM-6 and CM-10a) from
 RT Naja haie haie (Egyptian cobra) venom."
 RL Biochim. Biophys. Acta 537:1-8(1978).
 CC -1- MISCELLANEOUS: LD(50) IS 0.08 MG/KG BY INTRAVENOUS INJECTION.
 DR HSSP: P01426; 1NEA.
 DR InterPro: IPR003571; Snake_toxin.
 DR Pfam: PF00087; toxin: 1.
 DR ProDom: PD000206; Snake_toxin: 1.
 DR ProSITE: PS00272; SNAKE_TOXIN: 1.
 KW Venom; Neurotoxin; Postsynaptic neurotoxin; Multigene family.
 FT DISULFID 3 23 BY SIMILARITY.
 FT DISULFID 17 40 BY SIMILARITY.
 FT DISULFID 42 53 BY SIMILARITY.
 FT DISULFID 54 59 BY SIMILARITY.
 SO SEQUENCE 61 AA: 6888 MW: B22172627E93C0F0 CRC64;

Query Match 84.8%; Score 28; DB 1; Length 61;
 Best Local Similarity 66.7%; Pred. No. 23;
 Matches 4: Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKKXMR 6
 : 11 11
 Db 24 YKKWVR 29

RESULT 14
 NX52_NAJNI STANDARD: PRT: 61 AA.
 AC P01423:
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-MAY-1992 (Rel. 22, Last annotation update)
 DE SHORT NEUROTOXIN 2 (NEUROTOXIN BETA).
 OS Naja nivea (Cape cobra).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
 OC Elapidae; Elapinae; Naja.

OX NCBI_TaxID=8655;
 RN 11
 RP SEQUENCE.
 RC TISSUE-Venom;
 RX MEDLINE=72068724; PubMed=5167022;
 RA Botes D.P.;
 RT "Snake venom toxins. The amino acid sequences of toxins alpha and beta from Naja naja venom and the disulfide bonds of toxin alpha.";
 RL J. Biol. Chem. 246:7383-7391(1971).
 CC -1- FUNCTION: PRODUCES PERIPHERAL PARALYSIS BY BLOCKING NEUROMUSCULAR TRANSMISSION AT THE POSTSYNAPTIC SITE. BINDS TO THE NICOTINIC ACETYLCHOLINE RECEPTOR.
 CC -1- MISCELLANEOUS: LD(50) IS 0.08 MG/KG BY INTRAVENOUS INJECTION.
 CC PIR: A01692; NINJIC.
 DR HSSP; P01426; INEA.
 DR InterPro: IPR003571; Snake_toxin.
 DR Pfam: PF00087; toxin; 1.
 DR ProDom: PD000206; Snake_toxin; 1.
 DR PROSITE: PS00272; SNAKE_TOXIN; 1.
 KW Venom; Neurotoxin; Postsynaptic neurotoxin; Multigene family.
 FT DISULFID 3 23 BY SIMILARITY.
 FT DISULFID 17 40 BY SIMILARITY.
 FT DISULFID 42 53 BY SIMILARITY.
 FT DISULFID 54 59 BY SIMILARITY.
 SO SEQUENCE 61 AA: 6975 MW: 821731279B20F0 CRC64;

Query Match 84.8%; Score 28; DB 1; Length 61;
 Best Local Similarity 66.7%; Pred. No. 23;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKKXWR 6
 : 1 1 1
 Db 24 YKKRWR 29

RESULT 15
 NXS1_NAJKA STANDARD; PRT; 62 AA.
 ID NXS1_NAJKA
 AC P14613;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-MAY-1992 (Rel. 22, Last annotation update)
 DE SHORT NEUROTOXIN 1 (TOXIN C-6).
 OS Naja naja kaouthia (Monocled cobra) (Naja naja siamensis).
 OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi; Lepidodonta; Squamata; Scleroglossa; Serpentes; Colubroidea; Elapidae; Elapinae; Naja.
 OC NCBI_TaxID=8649;
 OX 11
 RN 11
 RP SEQUENCE.
 RC TISSUE-Venom;
 RX MEDLINE=90035789; PubMed=2807733;
 RA Chlou S.-H., Lin W.-W., Chang W.-P.;
 RT "Sequence characterization of venom toxins from Thailand cobra.";
 RL Int. J. Pept. Protein Res. 34:148-152(1989).
 CC -1- FUNCTION: PRODUCES PERIPHERAL PARALYSIS BY BLOCKING NEUROMUSCULAR TRANSMISSION AT THE POSTSYNAPTIC SITE. BINDS TO THE NICOTINIC ACETYLCHOLINE RECEPTOR.
 CC PIR: JK0221; JK0221.
 DR HSSP; P01430; ICOE.
 DR InterPro: IPR003571; Snake_toxin.
 DR Pfam: PF00087; toxin; 1.
 DR ProDom: PD000206; Snake_toxin; 1.
 DR PROSITE: PS00272; SNAKE_TOXIN; 1.
 KW Venom; Neurotoxin; Postsynaptic neurotoxin; Multigene family.
 FT DISULFID 3 24 BY SIMILARITY.
 FT DISULFID 17 41 BY SIMILARITY.
 FT DISULFID 43 54 BY SIMILARITY.
 FT DISULFID 55 60 BY SIMILARITY.
 SO SEQUENCE 62 AA: 6983 MW: 30975DA04B13E203 CRC64;

Query Match 84.8%; Score 28; DB 1; Length 62;
 Best Local Similarity 66.7%; Pred. No. 24;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKKXWR 6
 : 1 1 1
 Db 25 YKKRWR 30

Search completed: February 27, 2002, 11:52:16
 Job time: 563 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:50:20 ; Search time 62.61 Seconds
(without alignments)
14.017 Million cell updates/sec

Title: US-09-446-109a-21

Perfect score: .33

Sequence: 1 FRKXMR 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 14627329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTRMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	32	97.0	174 10 Q41626	Q41626 triticum ae
2	32	97.0	391 10 Q41538	Q41538 triticum ae
3	32	97.0	397 10 Q41536	Q41536 triticum ae
4	32	97.0	500 10 Q41537	Q41537 triticum ae
5	31	93.9	183 2 Q47491	Q47491 escherichia
6	31	93.9	209 2 Q31515	Q31515 bacillus su
7	31	93.9	348 5 Q21844	Q21844 caenorhabdi
8	31	93.9	377 2 Q9R085	Q9R085 delinococcus
9	31	93.9	1750 2 Q9RFX5	Q9RFX5 sligmatella
10	31	93.9	2207 5 Q22275	Q22275 caenorhabdi
11	30	90.9	382 4 Q9UH38	Q9UH38 homo sapien
12	29	87.9	62 13 Q91138	Q91138 najia najia (
13	29	87.9	62 13 Q91139	Q91139 najia najia (
14	29	87.9	83 13 Q9YGB6	Q9YGB6 najia sputat
15	29	87.9	83 13 Q9YGB5	Q9YGB5 najia sputat
16	29	87.9	88 5 Q9NN12	Q9NN12 leishmania
17	29	87.9	146 5 Q9A617	Q9A617 caulobacter
18	29	87.9	252 2 Q9P853	Q9P853 xylella fas
19	29	87.9	403 5 Q9GXM7	Q9GXM7 leishmania

20	29	87.9	428 5 Q9BWM9	Q9BWM9 trypanosoma
21	29	87.9	431 10 Q23147	Q23147 arabidopsis
22	29	87.9	433 10 Q42399	Q42399 brassica na
23	29	87.9	433 10 Q42636	Q42636 brassica na
24	29	87.9	726 2 Q9PFI5	Q9PFI5 xylella fas
25	29	87.9	726 2 Q9PAJ1	Q9PAJ1 xylella fas
26	29	87.9	788 12 Q9TMG8	Q9TMG8 lymantria d
27	29	87.9	893 10 Q9SWE6	Q9SWE6 hordium vul
28	29	87.9	893 10 Q9PSN6	Q9PSN6 najia najia (
29	29	87.9	80 5 Q23140	Q23140 caenorhabdi
30	28	84.8	82 13 Q9W6X0	Q9W6X0 najia atra (
31	28	84.8	83 13 Q57326	Q57326 najia sputat
32	28	84.8	83 13 Q57327	Q57327 najia sputat
33	28	84.8	83 13 Q9PFT0	Q9PFT0 najia najia (
34	28	84.8	83 13 Q9DE57	Q9DE57 najia atra (
35	28	84.8	87 13 Q9YGI9	Q9YGI9 bungarus mu
36	28	84.8	122 1 Q58942	Q58942 methanococ
37	28	84.8	198 1 Q9UYD1	Q9UYD1 pyrococcus
38	28	84.8	236 5 P91474	P91474 caenorhabdi
39	28	84.8	262 5 P91475	P91475 caenorhabdi
40	28	84.8	278 13 Q9PSH6	Q9PSH6 xenopus. dl
41	28	84.8	309 5 Q20847	Q20847 caenorhabdi
42	28	84.8	312 2 Q91522	Q91522 pseudomonas
43	28	84.8	383 2 Q51644	Q51644 enterococcu
44	28	84.8	384 2 Q52189	Q52189 enterococcu
45	28	84.8	384 2 Q52194	Q52194 enterococcu

ALIGNMENTS

RESULT 1
Q41626 PRELIMINARY; PRT: 174 AA.
AC Q41626:
BT 01-NOV-1996 (TREMBLREL. 01, Created)
DT 01-NOV-1996 (TREMBLREL. 01, Last sequence update)
DE 01-JUN-2001 (TREMBLREL. 17, Last annotation update)
DE LEUCINE RICH REPEAT REGION MRNA (FRAGMENT).
CS Triticum aestivum (Wheat).
CC Eukaryota: Viridiplantae; Streptophyta: Embryophyta: Tracheophyta:
OC Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae: Poideae:
OC Triticaceae: Triticum.
CX NCBI_TaxID=4565;
KN [1]
RP SEQUENCE FROM N.A.
KC STRAIN=CV. CHINESE SPRING:
RX MEDLINE=94217352; PubMed=8190067;
KA J1 L.H., Langridge P.;
BT "An early mitosis cDNA clone from wheat."
FT MOL. Gen. Evol. 243:17-23(1994).
DR EMBL: X79130; CA55731.1; ..
DR Mendel: 24275; Tritic; 2361; 24276.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR003592; LRR_out.
DR Pfam: PF00560; LRR_3.
DR PRINTS: PR00019; LEURICRPT.
DR SMART: SM00370; LRR_3.
FT NON_TER
SO SEQUENCE 174 AA: 18919 MW: PC2B1344ADDC0633 CRC64:

Query Match 97.0%; Score 32; DB 10; Length 174;

Best Local Similarity 83.3%; Pred. No. 50;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FRKXMR 6
DB 149 FRKXMR 154

RESULT 2
Q41538

ID Q41538 PRELIMINARY: PRT: 391 AA.
 AC Q41538;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE (SUBCLONE PAMJL236) AWJL236 GENE.
 GN AWJL236.
 OS Triticum aestivum (wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticaceae; Triticum.
 OC NCBI_TaxID=4565;
 RX InterPro: IPR003592; LRR_out.
 RP STRAIN-CV. CHINESE SPRING;
 RA Langhui J., Letarte J., Langridge P.;
 RL Submitted (SEP-1994) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. CHINESE SPRING;
 RA JI L.;
 RL Submitted (SEP-1994) to the EMBL/Genbank/DBJ databases.
 DE EMBL: X81370; CA57135.1; -;
 DR Mendel: 14338; Triae:2361;14338.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR003592; LRR_out.
 DR Pfam: PF00560; LRR: 4.
 DR SMART: SM00370; LRR: 8.
 SO SEQUENCE 391 AA: 42810 MW: 16F0779762DDC1BC5 CMC64;

Query Match 97.0%; Score 32; DB 10; Length 391;
 Best Local Similarity 83.3%; Pred. No. 1.1e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FKKXMR 6
 ||| ||
 Db 350 FKKTWR 355

RESULT 3
 Q41536 PRELIMINARY: PRT: 397 AA.
 ID Q41536;
 AC Q41536;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE (SUBCLONE PAMJL175) AWJL175 GENE.
 GN AWJL175.
 OS Triticum aestivum (wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticaceae; Triticum.
 OC NCBI_TaxID=4565;
 RX InterPro: IPR001611; LRR.
 RP STRAIN-CV. CHINESE SPRING;
 RA Langhui J., Letarte J., Langridge P.;
 RL Submitted (SEP-1994) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. CHINESE SPRING;
 RA JI L.;
 RL Submitted (SEP-1994) to the EMBL/Genbank/DBJ databases.
 DE EMBL: X81368; CA57133.1; -;
 DR Mendel: 14336; Triae:2359;14336.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR003592; LRR_out.
 DR Pfam: PF00560; LRR: 6.
 DR PRINTS: PR00019; LEURICHRPT.
 DR SMART: SM00370; LRR: 9.
 SO SEQUENCE 397 AA: 44215 MW: 20HECCFUD3CDBER9 CMC64;

Query Match 97.0%; Score 32; DB 10; Length 397;
 Best Local Similarity 83.3%; Pred. No. 1.1e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FKKXMR 6
 ||| ||
 Db 360 FKKTWR 365

RESULT 4
 Q41537 PRELIMINARY: PRT: 500 AA.
 ID Q41537;
 AC Q41537;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE (SUBCLONE PAMJL218) AWJL218 GENE.
 GN AWJL218.
 OS Triticum aestivum (wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticaceae; Triticum.
 OC NCBI_TaxID=4565;
 RX InterPro: IPR001611; LRR.
 RP STRAIN-CV. CHINESE SPRING;
 RA Langhui J., Letarte J., Langridge P.;
 RL Submitted (SEP-1994) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. CHINESE SPRING;
 RA JI L.;
 RL Submitted (SEP-1994) to the EMBL/Genbank/DBJ databases.
 DE EMBL: X81369; CA57134.1; -;
 DR Mendel: 14337; Triae:2360;14337.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR003592; LRR_out.
 DR Pfam: PF00560; LRR: 8.
 DR PRINTS: PR00019; LEURICHRPT.
 DR SMART: SM00370; LRR: 10.
 SO SEQUENCE 500 AA: 56042 MW: C29700F226DE92EA CMC64;

Query Match 97.0%; Score 32; DB 10; Length 500;
 Best Local Similarity 83.3%; Pred. No. 1.3e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FKKXMR 6
 ||| ||
 Db 463 FKKTWR 468

RESULT 5
 Q47491 PRELIMINARY: PRT: 183 AA.
 ID Q47491;
 AC Q47491;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
 DE PUTATIVE P1FA PROTEIN (FRAGMENT).
 GN P1FA.
 OS Escherichia coli.
 OC Plasmid mini-F.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OC NCBI_TaxID=562;
 RX InterPro: IPR001611; LRR.
 RP SEQUENCE FROM N.A.
 RA MEDLINE=87117528; PubMed=3027660;
 RA Caughey P.A., De Feyter R., Lane H.E.D.;
 PT "The minif plasmid C protein: sequence, purification and DNA
 binding.";

RL Nucleic Acids Res. 14:9699-9712(1986).

DR EMBL: X04968; CAA28642.1; -

KM plasmid.

FT NON_TER 183 183

SO SEQUENCE 183 AA; 20396 MW; 7771C907DCA0B341 CRC64;

Query Match 93.9%; Score 31; DB 2; Length 183;

Best Local Similarity 83.3%; Pred. NO. 81;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FKKXMR 6

DB 117 FKKLMR 122

RESULT 6 PRELIMINARY; PRT; 209 AA.

ID 031515

AC 031515

DT 01-JAN-1998 (TEMBLrel. 05, Created)

DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)

DT 01-NOV-1998 (TEMBLrel. 08, Last annotation update)

DE YESL PROTEIN.

OS Bacillus subtilis.

OC Bacteria: Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

NCBI_TaxID=1423;

SEQUENCE FROM N.A.

RC STRAIN=168;

RE MEDLINE=98044033; PubMed=9384377;

RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,

RA Azevedo V., Bertolo M.C., Bessieres P., Bolotin A., Borchert S.,

RA Boursier R., Boursier L., Brans A., Braun M., Brintnell S.C., Bron S.,

RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,

RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,

RA Dantot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,

RA Ertlen K.D., Errington J., Fabret C., Ferrari E., Foulger D.,

RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,

RA Gilm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,

RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,

RA Joris R., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,

RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,

RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,

RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,

RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,

RA Noone D., O'Reilly M., Ogawa K., Ogilwara A., Oudega B., Park S.H.,

RA Prescan E., Fujic P., Purnelle D., Porwollik S., Prescott A.M.,

RA Rieger M., Rivoita C., Roche E., Roche B., Rose M., Sadale Y.,

RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,

RA Sekiuchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,

RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,

RA Takeuchi M., Tamakoshi A., Tanaka T., Terstra P., Tognoni A.,

RA Tosato V., Uchiyama S., Vandendol M., Vannier F., Vassarotti A.,

RA Viart A., Wambuit R., Wedler E., Wedler H., Weltenegeger T.,

RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,

RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa K., Danchin A.,

RT *The complete genome sequence of the gram-positive bacterium Bacillus

subtilis.

RL Nature 390:249-256(1997).

RN 121

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;

RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL: Z99107; CAB12513.1; -

KM Complete proteome.

SO SEQUENCE 209 AA; 23436 MW; 80FABCAAC1690335 CRC64;

Query Match 93.9%; Score 31; DB 2; Length 209;

Best Local Similarity 83.3%; Pred. NO. 92;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FKKXMR 6

DB 72 FKKEMR 77

RESULT 7

ID 021844

AC 021844

DT 01-NOV-1996 (TEMBLrel. 01, Created)

DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)

DT 01-NOV-1998 (TEMBLrel. 08, Last annotation update)

DE COSMID R08C7.

OS Caenorhabditis elegans.

OC Eukaryota: Metazoa: Nematoda; Chromadorea; Rhabditida; Rhabditidae;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

NCBI_TaxID=6239;

SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RE MEDLINE=94150718; PubMed=7906398;

RA Bonfield J., Ainscough R., Anderson K., Baynes C., Berks M.,

RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,

RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,

RA Jones M., Kershaw J., Kirsten J., Lalster N., Latreille P.,

RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,

RA Smalton D., Smith A., Sonnenhammer E., Staden R., Sulston J.,

RA Telford-Wieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,

RA Watson A., Weinstock L., Wilkinson-Spratt J., Wohlman P.,

RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

elegans.

RL Nature 368:32-38(1994).

RN 121

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Favello A., Le T.T.;

RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL: U61953; AAC48073.1; -

SO SEQUENCE 348 AA; 38572 MW; A1BB018706407492 CRC64;

Query Match 93.9%; Score 31; DB 5; Length 348;

Best Local Similarity 83.3%; Pred. NO. 1,5e+02;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FKKXMR 6

DB 250 FKKLMR 255

RESULT 8

ID 09R085

AC 09R085

DT 01-MAY-2000 (TEMBLrel. 13, Created)

DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)

DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)

DE CONSERVED HYPOTHETICAL PROTEIN.

OS Deinococcus radiodurans.

OC Bacteria: Thermus/Deinococcus group; Deinococcales; Deinococcus.

NCBI_TaxID=1299;

KN 11

RP SEQUENCE FROM N.A.

RC STRAIN=RI;

RX MEDLINE=20036896; PubMed=10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Halt D.H., Gwin M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioresistant bacterium *Deinococcus*
 RT *radiodurans* RI.";
 RL Science 286:1571-1577(1999).
 DR EMBL: AE001994; AAF11073.1; -.
 DR TIGR: DR1507; -.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR000130; Zn_MpPtdase.
 DR Pfam: PF00595; PDZ. 1.
 DR SMART: SM00228; PDZ. 1.
 DR PROSITE: PSS0106; PDZ. 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.
 KM Complete proteome.
 SQ SEQUENCE 377 AA; 40486 MW; 58553505806DC64C CRC64;

Query Match 93.9%; Score 31; DB 2; Length 377;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKKXMR 6
 DB 57 FKKLMR 62

RESULT 9
 ID 09RFK5 PRELIMINARY; PRT: 1750 AA.
 AC 09RFK5;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE MTAG.
 GN MTAG.
 OS *Stigmatella aurantiaca*.
 OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
 OC Myxococcales; Cystobacterineae; Cystobacteraceae; *Stigmatella*.
 OX NCBI_TaxID=41;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=DM4/3-1;
 RX MEDLINE=20069734; PubMed=10601310;
 RA Sliakowski B., Schaller H.U., Ehret H., Kunze B., Weidig S.,
 RA Nordstiek G., Brandt P., Bloeker H., Hoeffle G., Beyer S., Mueller R.;
 RT "New Lessons for Combinatorial Biosynthesis from Myxobacteria: The
 RT Myxothiazol Biosynthetic Gene Cluster of *Stigmatella aurantiaca* DM4/3-
 RT 1.";
 RL J. Biol. Chem. 274:37391-37399(1999).
 DR EMBL: AF188287; AAF19815.1; -.
 DR HSSP: P14687; IAMU.
 DR InterPro: IPR000873; AMP-Bind.
 DR InterPro: IPR002103; Bac_Luciferase.
 DR InterPro: IPR001242; DUF4.
 DR InterPro: IPR003880; Phosphopant_attach.
 DR InterPro: IPR001031; Thioesterase.
 DR Pfam: PF00501; AMP-binding. 1.
 DR Pfam: PF00296; bac_luciferase. 1.
 DR Pfam: PF00668; Condensation. 1.
 DR Pfam: PF00550; pp-binding. 1.
 DR Pfam: PF00975; Thioesterase. 1.
 DR PRINTS: PR00154; AMPBINDING.
 DR PROSITE: PSS0075; ACP_DOMAIN. 1.
 KW Phosphopantetheine.
 SQ SEQUENCE 1750 AA; 191776 MW; A9D947C08F730CCA CRC64;

Query Match 93.9%; Score 31; DB 2; Length 1750;
 Best Local Similarity 83.3%; Pred. No. 6.9e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKKXMR 6
 DB 886 FKKLMR 891

RESULT 10
 ID 022275 PRELIMINARY; PRT: 2207 AA.
 AC 022275; 023214;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE HYPOTHEICAL 244.6 KDA PROTEIN W07E11.1 IN CHROMOSOME X.
 GN W07E11.1
 OS *Caenorhabditis elegans*.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; *Caenorhabditis*.
 OX NCBI_TaxID=6239;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Kershaw J.;
 HL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- COFACTOR: FAD (BY SIMILARITY).
 CC -1- SIMILARITY: TO PYRIDINE NUCLEOTIDE-DISULPHIDE OXIDOREDUCTASES
 CC CLASS-I.
 DR EMBL: Z49889; CA90070.1; -.
 DR EMBL: Z49868; CA90070.1; JOINED.
 DR EMBL: Z49868; CA90032.1; -.
 DR EMBL: Z49889; CA90032.1; JOINED.
 DR WormPep: W07E11.1; CE023372.
 DR InterPro: IPR000759; Adrxn_reductase.
 DR InterPro: IPR002489; DUF14.
 DR InterPro: IPR001327; FAD_pyr_redox.
 DR InterPro: IPR002932; Glu_synchase.
 DR InterPro: IPR000205; NAD_binding.
 DR InterPro: IPR000103; pyridine_redox_2.
 DR InterPro: IPR001100; pyr_redox.
 DR Pfam: PF01493; DUF14. 1.
 DR Pfam: PF01645; Glu_synchase. 1.
 DR PRINTS: PR00419; ADXRDFASE.
 DR PRINTS: PR00368; FADPNR.
 DR PRINTS: PR00411; PNDROFASEI.
 DR PRINTS: PR00469; PNDROFASEII.
 KW FAD; Flavoprotein; hypochlorite protein; Oxidoreductase;
 KW Redox-active center.
 SQ SEQUENCE 2207 AA; 244635 MW; 8FDA8462AC287D1A CRC64;

Query Match 93.9%; Score 31; DB 5; Length 2207;
 Best Local Similarity 83.3%; Pred. No. 8.6e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKKXMR 6
 DB 1775 FKKLMR 1780

RESULT 11
 ID 09JHJ8 PRELIMINARY; PRT: 392 AA.
 AC 09JHJ8;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE ARYL SULFATASE D BETA.
 GN ARSD.
 OS *Homo sapiens* (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo:
 OX NCBI_TaxID=9606;
 RN 11
 RP SEQUENCE FROM N.A.
 RC TISSUE=TESTIS;
 RA Uchitsch P., Salzer M.J., Mirschmann P., Vogt P.H.;
 RT "The arylsulfatase D (ARSD) gene in Xp22.3 encodes two protein
 RL isoforms competing in vitro with arylsulfatase C proteins."
 DR Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF160499; AAF22253.1;
 DR HSSP: P15848; IFSU.
 DR InterPro: IPR000917; Sulfatase.
 DR Pfam: PF00884; Sulfatase.
 DR PROSITE: PS00523; SULFATASE_1; 1.
 DR PROSITE: PS00149; SULFATASE_2; 1.
 SO SEQUENCE 382 AA; 42310 MW; 294954BEACCTF46D CRC64;

Query Match
 Best Local Similarity 90.9%; Score 30; DB 4; Length 382;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 FKXMR 6
 Db 366 FKXMR 371

RESULT 12
 ID 091138 PRELIMINARY; PRT; 62 AA.
 AC 091138;
 DT 01-NOV-1996 (TREMURel. 01, Created)
 DT 01-NOV-1996 (TREMURel. 01, Last sequence update)
 DE 01-JUN-2001 (TREMURel. 17, Last annotation update)
 OS NEUROTOXIN (FRAGMENT)
 OS Naja naja (Indian cobra).
 OC Eukaryota: Metazoa: Chordata: Cranialata: Vertebrata: Euteleostomi:
 OC Lepidodactylus: Squamata: Scleroglossa: Serpentes: Colubroidae;
 OC Elapidae: Elapinae: Naja.
 NCBI_TaxID=35670;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPUTATRIX; TISSUE=VENOM GLAND;
 RA Jeyaseelan K., Armugam A., Lachumanan R., Earnest L., Tan N.H.,
 RA Tan C.H., Gopalakrishnakone P.P.;
 RL Submitted (MAY-1995) to the EMBL/Genbank/DBJ databases.
 DR EMBL: L42003; AAA66025.1;
 DR HSSP: P01430; ICOD.
 DR InterPro: IPR003571; Snake_toxin.
 DR Pfam: PF00087; toxin; 1.
 DR PRODOM: PD000206; Snake_toxin; 1.
 DR PROSITE: PS00272; SNAKE_TOXIN; 1.
 KW Neurotoxin.
 FT NON_TER
 SO SEQUENCE 62 AA; 6916 MW; C39B87CB40CEEBFF CRC64;

Query Match
 Best Local Similarity 87.9%; Score 29; DB 13; Length 62;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 FKXMR 6
 Db 25 YKXMR 30

RESULT 13
 ID 091139 PRELIMINARY; PRT; 62 AA.
 AC 091139;
 DT 01-NOV-1996 (TREMURel. 01, Created)
 DT 01-NOV-1996 (TREMURel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMURel. 17, Last annotation update)

DE NEUROTOXIN (FRAGMENT)
 CS Naja naja (Indian cobra).
 OC Eukaryota: Metazoa: Chordata: Cranialata: Vertebrata: Euteleostomi;
 OC Lepidodactylus: Squamata: Scleroglossa: Serpentes: Colubroidae;
 OC Elapidae: Elapinae: Naja.
 NCBI_TaxID=35670;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPUTATRIX; TISSUE=VENOM GLAND;
 RA Jeyaseelan K., Armugam A., Lachumanan R., Earnest L., Tan N.H.,
 RA Tan C.H., Gopalakrishnakone P.P.;
 RL Submitted (MAY-1995) to the EMBL/Genbank/DBJ databases.
 DR EMBL: L42003; AAA66026.1;
 DR HSSP: P01430; ICOD.
 DR InterPro: IPR003571; Snake_toxin.
 DR Pfam: PF00087; toxin; 1.
 DR PRODOM: PD000206; Snake_toxin; 1.
 DR PROSITE: PS00272; SNAKE_TOXIN; 1.
 KW Neurotoxin.
 FT NON_TER
 SO SEQUENCE 62 AA; 6885 MW; 779B87D1EBB82A39 CRC64;

Query Match
 Best Local Similarity 87.9%; Score 29; DB 13; Length 62;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 FKXMR 6
 Db 25 YKXMR 30

RESULT 14
 ID 09YJG6 PRELIMINARY; PRT; 83 AA.
 AC 09YJG6;
 DT 01-MAY-1999 (TREMURel. 10, Created)
 DT 01-MAY-1999 (TREMURel. 10, Last sequence update)
 DE 01-JUN-2001 (TREMURel. 17, Last annotation update)
 OS POST SYNAPTIC ALPHA NEUROTOXIN PRECURSOR.
 OS Naja sputatrix (Malayan spitting cobra).
 OC Eukaryota: Metazoa: Chordata: Cranialata: Vertebrata: Euteleostomi;
 OC Lepidodactylus: Squamata: Scleroglossa: Serpentes: Colubroidae;
 OC Elapidae: Elapinae: Naja.
 NCBI_TaxID=33626;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPUTATRIX; TISSUE=LIVER;
 RA Afiliyan F., Lim C.Y., Armugam A., Tan C.H., Gopalakrishnakone P.,
 RA Jeyaseelan K.;
 RT "Post synaptic alpha neurotoxin gene of the spitting cobra, Naja sputatrix: Structure, organization and phylogenetic analysis."
 KL Genome Res. 0:0-0(1999).
 DR EMBL: AF096999; AAD08812.1;
 DR HSSP: P01430; ICOD.
 DR InterPro: IPR003571; Snake_toxin.
 DR Pfam: PF00087; toxin; 1.
 DR PRODOM: PD000206; Snake_toxin; 1.
 DR PROSITE: PS00272; SNAKE_TOXIN; 1.
 KW Signal: Neurotoxin.
 FT SIGNAL
 FT CHAIN
 SO SEQUENCE 83 AA; 9219 MW; F9CAD71A93B7AE75 CRC64;

Query Match
 Best Local Similarity 87.9%; Score 29; DB 13; Length 83;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 FKXMR 6
 Db 46 YKXMR 51

RESULT 15

09YGCJ5 PRELIMINARY; PRT: 83 AA.

AC 09YGCJ5: 01-MAY-1999 (TREMBLrel. 10, Created)

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE POST SYNAPTIC ALPHA NEUROTOXIN PRECURSOR.

CN MTX.

OS Naja sputatrix (Malayan spitting cobra).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Lepidosaurs; Squamata; Scleroglossa; Serpentes; Colubroidea;

OC Elapidae; Elapinae; Naja.

OX NCBI_TaxID=33626;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=SPUTATRIX; TISSUE=LIVER;

RA Alfliyan F., Lim C.Y., Armugam A., Tan C.H., Gopalakrishnakone P.,

RA Jayaseelan K.;

RT "Post synaptic alpha neurotoxin gene of the spitting cobra, Naja naja

RT sputatrix: Structure, organization and phylogenetic analysis.";

RL Genome Res. 0:0-0(1999). 1: -.

DR EMBL: AF097000; AAD08813.1: -.

DR HSSP: P01430; ICOD.

DR InterPro: IPR003571; Snake_toxin.

DR Pfam: PF00087; toxin; 1.

DR ProDom: PD000206; Snake_toxin; 1.

DR PROSITE: PS00272; SNAKE_TOXIN; 1.

KW Signal; Neurotoxin.

FT SIGNAL 1 21 POTENTIAL.

FT CHAIN 22 83 POST SYNAPTIC ALPHA NEUROTOXIN.

SO SEQUENCE 83 AA; 9188 MW; 4DCAD70038C16EB3 CRC64;

Query Match 87.9%; Score 29; DB 13; Length 83;

Best Local Similarity 66.7%; Pred. No. 92;

Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKXWR 6

DB 46 YKXWR 51

Search completed: February 27, 2002, 11:54:08

Job time: 228 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:36:11 : Search time 32.48 Seconds
(without alignments)
4.157 Million cell updates/sec

Title: US-09-446-109A-21

Perfect score: 33

Sequence: 1 FKXMR 6

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: 1: /cgn2_6/prodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/prodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	97.0	37	1	US-08-665-543B-4
2	29	87.9	386	6	Sequence 4, Appl 1
3	29	87.9	386	6	Patent No. 5171840-5
4	29	87.9	433	3	Patent No. 5480796-5
5	29	87.9	468	4	Sequence 6, Appl 1
6	29	87.9	468	6	Sequence 5, Appl 1
7	29	87.9	468	6	Patent No. 5171840-2
8	28	84.8	17	1	Patent No. 5480796-2
9	28	84.8	17	1	Sequence 9, Appl 1
10	27	81.8	52	4	Sequence 6, Appl 1
11	27	81.8	52	4	Sequence 17, Appl 1
12	27	81.8	123	1	Sequence 15, Appl 1
13	27	81.8	123	1	Sequence 20, Appl 1
14	27	81.8	125	1	Sequence 20, Appl 1
15	27	81.8	125	4	Sequence 20, Appl 1
16	27	81.8	148	3	Sequence 20, Appl 1
17	27	81.8	160	2	Sequence 12, Appl 1
18	27	81.8	213	1	Sequence 14, Appl 1
19	27	81.8	213	1	Sequence 23, Appl 1
20	27	81.8	215	1	Sequence 22, Appl 1
21	27	81.8	215	4	Sequence 23, Appl 1
22	27	81.8	218	1	Sequence 23, Appl 1
23	27	81.8	218	2	Sequence 23, Appl 1
24	27	81.8	234	1	Sequence 4, Appl 1
25	27	81.8	234	1	Sequence 8, Appl 1
26	27	81.8	234	1	Sequence 8, Appl 1
27	27	81.8	234	1	Sequence 8, Appl 1
27	27	81.8	234	4	Sequence 8, Appl 1

28	27	81.8	791	1	US-08-394-880B-2	Sequence 2, Appl 1
29	27	81.8	2183	3	US-08-746-111-5	Sequence 5, Appl 1
30	26	78.8	6	2	US-08-637-759B-179	Sequence 179, Appl 1
31	26	78.8	6	3	US-08-871-355A-179	Sequence 179, Appl 1
32	26	78.8	29	3	US-08-374-077C-19	Sequence 61, Appl 1
33	26	78.8	29	4	US-08-895-590-61	Sequence 4, Appl 1
34	26	78.8	332	3	US-09-183-253-4	Sequence 1, Appl 1
35	26	78.8	411	1	US-08-232-532-1	Sequence 1, Appl 1
36	26	78.8	411	2	US-08-748-150-1	Sequence 1, Appl 1
37	26	78.8	411	5	US-09-347-877-1	Sequence 1, Appl 1
38	26	78.8	416	3	US-09-320-878-18	Sequence 1, Appl 1
39	26	78.8	416	4	US-09-105-537-39	Sequence 18, Appl 1
40	26	78.8	453	3	US-08-206-176-6	Sequence 39, Appl 1
41	26	78.8	454	3	US-08-105-537-39	Sequence 27, Appl 1
42	26	78.8	454	3	US-08-434-099A-27	Sequence 28, Appl 1
43	26	78.8	582	4	US-08-434-099A-28	Sequence 1, Appl 1
44	26	78.8	582	4	US-08-448-489-1	Sequence 1, Appl 1
45	26	78.8	622	1	US-08-547-197-1	Sequence 1, Appl 1

ALIGNMENTS

RESULT 1
US-08-665-543B-4
Sequence 4, Application US/08665543B
Patent No. 5734015
GENERAL INFORMATION:
APPLICANT: Shinnar, Ann
APPLICANT: Zaslloff, Michael A.
TITLE OF INVENTION: NEW FAMILY OF LINEAR ANTIMICROBIAL
TITLE OF INVENTION: PEPTIDES FROM HAGRISH INTESTINE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESS: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/665,543B
FILING DATE: 18-JUN-1996
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/000,325
FILING DATE: 19-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Barlow, Stacey A.
REGISTRATION NUMBER: 39,595
REFERENCE/DOCKET NUMBER: 05387.0036-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4000
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: Linear
MOLECULE TYPE: peptide
US-08-665-543B-4
Query Match: 97.0%; Score 32; DB 1; Length 37;
Best Local Similarity 83.3%; Pred. No. 6.3;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKKXMR 6
111 11

Db 3 FKKAMR 8

RESULT 2

5171840-5

Patent No. 5171840

APPLICANT: KISHIMOTO, TADAMITSU

TITLE OF INVENTION: RECEPTOR PROTEIN FOR HUMAN B CELL

STIMULATORY FACTOR-2

NUMBER OF SEQUENCES: 11

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/298,694

FILING DATE: 19-JAN-1989

SEQ ID NO: 5:

LENGTH: 386

5171840-5

Query Match. 87.9%; Score 29; DB 6; Length 386;

Best Local Similarity 66.7%; Pred. No. 2e+02;

Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKKXMR 6
111 11

Db 306 FKKTKW 311

RESULT 3

5480796-5

Patent No. 5480796

APPLICANT: KISHIMOTO, TADAMITSU

TITLE OF INVENTION: ANTIBODIES AGAINST THE RECEPTOR PROTEIN

FOR HUMAN B CELL STIMULATORY FACTOR-2

NUMBER OF SEQUENCES: 8

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/907,650

FILING DATE: 02-JUL-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 298,694

FILING DATE: 19-JAN-1989

SEQ ID NO: 5:

LENGTH: 386

5480796-5

Query Match. 87.9%; Score 29; DB 6; Length 386;

Best Local Similarity 66.7%; Pred. No. 2e+02;

Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKKXMR 6
111 11

Db 306 FKKTKW 311

RESULT 4

US-08-941-532-6

Sequence 6, Application US/08941532

Patent No. 6096946

GENERAL INFORMATION:

APPLICANT: ROBERTS, Jeremy Alan

APPLICANT: COUPE, Simon Allan

APPLICANT: JENKINS, Elizabeth Sarah

TITLE OF INVENTION: CONTROL OF POD DEHISCENCE

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox

STREET: 1100 New York Avenue

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30 (PRO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/941,532

FILING DATE: 30-SEP-1997

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/G896/00757

FILING DATE: 29-MAR-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9506684.1

FILING DATE: 31-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Esmond, Robert W.

REGISTRATION NUMBER: 32,893

REFERENCE/DOCKET NUMBER: 0623.0580001/RWE

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2600

TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 433 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-941-532-6

Query Match. 87.9%; Score 29; DB 3; Length 433;

Best Local Similarity 66.7%; Pred. No. 2.2e+02;

Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKKXMR 6
111 11

Db 88 FKKAMK 93

RESULT 5

US-08-795-473B-5

Sequence 5, Application US/08795473B

Patent No. 6217858

GENERAL INFORMATION:

APPLICANT: Galun, Elhan

APPLICANT: Nahot, Orit

APPLICANT: Blum, Herbert E.

TITLE OF INVENTION: A Pharmaceutical Composition for Treating

TITLE OF INVENTION: Hepatitis B Virus (HBV) Infection

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Davidson, Davidson and Kappel, LLC

STREET: 1140 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: MS-DOS EDITOR

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/795,473B

FILING DATE: 11-FEB-1997

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Davidson, Clifford M.

REGISTRATION NUMBER: 32,728

REFERENCE/DOCKET NUMBER: 963.1007

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)-997-1028
TELEFAX: (212)-997-1037
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 468 amino acids
TYPE: amino acid
TOPOLOGY: unknown
US-08-795-473B-5

Query Match 87.9%: Score 29; DB 6; Length 468;
Best Local Similarity 66.7%: Pred. No. 2.4e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKKXWR 6
DB 388 FKKTWK 393

RESULT 6
5171840-2
Patent No. 5171840
APPLICANT: KISHIMOTO, TADAMITSU
TITLE OF INVENTION: RECEPTOR PROTEIN FOR HUMAN B CELL
STIMULATORY FACTOR-2
NUMBER OF SEQUENCES: 11
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/298,694
FILING DATE: 19-JAN-1989
SEQ ID NO: 2
LENGTH: 468
5171840-2

Query Match 87.9%: Score 29; DB 6; Length 468;
Best Local Similarity 66.7%: Pred. No. 2.4e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKKXWR 6
DB 388 FKKTWK 393

RESULT 7
5480796-2
Patent No. 5480796
APPLICANT: KISHIMOTO, TADAMITSU
TITLE OF INVENTION: ANTIBODIES AGAINST THE RECEPTOR PROTEIN
FOR HUMAN B CELL STIMULATORY FACTOR-2
NUMBER OF SEQUENCES: 8
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/907,650
FILING DATE: 02-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 298,694
FILING DATE: 19-JAN-1989
SEQ ID NO: 2
LENGTH: 468
5480796-2

Query Match 87.9%: Score 29; DB 6; Length 468;
Best Local Similarity 66.7%: Pred. No. 2.4e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKKXWR 6
DB 388 FKKTWK 393

RESULT 8
US-08-213-124-9

Sequence 9, Application US/08213124
Patent No. 5693325

GENERAL INFORMATION:
APPLICANT: Kahn, Michael
TITLE OF INVENTION: PEPTIDE VACCINES AND METHODS RELATING
THERE TO
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/213,124
FILING DATE: 15-MAR-1994
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
NAME: Heilmann, Karl R.
REGISTRATION NUMBER: 33,507
REFERENCE/DOCKET NUMBER: 670063.411
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANDBERK

INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-213-124-9

Query Match 84.8%: Score 28; DB 1; Length 17;
Best Local Similarity 66.7%: Pred. No. 17;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKKXWR 6
DB 1 YKKVWR 6

RESULT 9
US-09-196-390-6
Sequence 6, Application US/09196390
Patent No. 6307125
GENERAL INFORMATION:
APPLICANT: K, Martina
APPLICANT: K, Horst
APPLICANT: Mullicks, Stephanie
APPLICANT: Walter, Lennart
APPLICANT: Froberg, Claus

APPLICANT: Kossmann, Jens
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING ENZYMES
TITLE OF INVENTION: FROM WHEAT WHICH ARE INVOLVED IN STARCH
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: James F. Haley, Jr., C/O Fish & Neave
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10020

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/196,390
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE 196 21 588.9
FILING DATE: 29-MAY-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE 196 36 917.7
FILING DATE: 11-SEP-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP97/02793
FILING DATE: 28-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Haley, Jr., James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: AGREVO-9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 596-9000
TELEFAX: (212) 596-9090
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 799 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-196-390-6

Query Match 84.8%; Score 28; DB 4; Length 799;
Best Local Similarity 66.7%; Pred. No. 5.9e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 FKKXW 6
DB 763 FKESW 768
RESULT 10
US-09-369-494-17
Sequence 17, Application US/09369494
Patent No. 6180607
GENERAL INFORMATION:
APPLICANT: Davies, Christopher
APPLICANT: Chen, Dadong
APPLICANT: Rocznik, Steve
TITLE OF INVENTION: Protein Having Proteinase Inhibitor Activity
FILE REFERENCE: MSB-7260
CURRENT APPLICATION NUMBER: US/09/369,494
CURRENT FILING DATE: 1999-08-05
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 17
LENGTH: 52
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: fragment from
US-09-369-494-17

Query Match 81.8%; Score 27; DB 4; Length 52;
Best Local Similarity 80.0%; Pred. No. 72;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 FKKXW 5
DB 4 FKXSW 8

RESULT 11
US-09-358-569D-15
Sequence 15, Application US/09358569D
Patent No. 6294648
GENERAL INFORMATION:
APPLICANT: Delaria, Kathy
APPLICANT: Rocznik, Steve
APPLICANT: Davies, Christopher
TITLE OF INVENTION: Protein Having Proteinase Inhibitor Activity
FILE REFERENCE: MSB-7259
CURRENT APPLICATION NUMBER: US/09/358,569D
CURRENT FILING DATE: 1999-07-20
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 15
LENGTH: 52
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: from computer
US-09-358-569D-15

Query Match 81.8%; Score 27; DB 4; Length 52;
Best Local Similarity 80.0%; Pred. No. 72;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 FKKXW 5
DB 4 FKXSW 8

RESULT 12
US-08-225-989-20
Sequence 20, Application US/08225989
Patent No. 5480981
GENERAL INFORMATION:
APPLICANT: Goodwin, Raymond G.
APPLICANT: Smith, Craig A.
APPLICANT: Armitage, Richard J.
APPLICANT: Gruss, Hans-Jurgen
TITLE OF INVENTION: NO. 5480981el Cytokine That Binds CD30
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kathryn A. Seese, Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple 7.1
SOFTWARE: Microsoft Word, Version 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/225,989
FILING DATE: 12 APRIL 1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/966,775
FILING DATE: 27-OCT-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 907,224
FILING DATE: 01-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 899,660
FILING DATE: 15-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 892,459

FILING DATE: 02-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 889,717
FILING DATE: 26-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Seese, Kathryn A.
REGISTRATION NUMBER: 32,172
REFERENCE/DOCKET NUMBER: 2804-E
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
TELEFAX: (206)233-0644
TELEX: 756822
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 125 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein fragment
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: hucd30 fragment (PRELIM)
US-08-225-989-20

Query Match 81.88; Score 27; DB 1; Length 125;
Best Local Similarity 80.08; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 1 FKKXW 5
|||
Db 81 FKKSW 85

RESULT 13
US-08-570-923-20
Sequence 20, Application US/08570923
Patent No. 5677430
GENERAL INFORMATION:
APPLICANT: Goodwin, Raymond G.
APPLICANT: Smith, Craig A.
APPLICANT: Armitage, Richard J.
APPLICANT: Gruss, Hans-Jurgen
TITLE OF INVENTION: No. 5677430el Cytokine That Binds CD30
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kathryn A. Seese, Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: Apple 7.1
SOFTWARE: Microsoft Word, Version 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/570,923
FILING DATE: 12-DEC-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/225,989
FILING DATE: 12 APRIL 1994
APPLICATION NUMBER: US 07/966,775
FILING DATE: 27-OCT-1992
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 907,224
FILING DATE: 01-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 899,660

FILING DATE: 15-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 892,459
FILING DATE: 02-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 889,717
FILING DATE: 26-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Seese, Kathryn A.
REGISTRATION NUMBER: 32,172
REFERENCE/DOCKET NUMBER: 2804-E
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
TELEFAX: (206)233-0644
TELEX: 756822
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 125 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein fragment
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: hucd30 fragment (PRELIM)
US-08-570-923-20

Query Match 81.88; Score 27; DB 1; Length 125;
Best Local Similarity 80.08; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 1 FKKXW 5
|||
Db 81 FKKSW 85

RESULT 14
US-08-580-014-20
Sequence 20, Application US/08580014
Patent No. 5753203
GENERAL INFORMATION:
APPLICANT: Goodwin, Raymond G.
APPLICANT: Smith, Craig A.
APPLICANT: Armitage, Richard J.
APPLICANT: Gruss, Hans-Jurgen
TITLE OF INVENTION: No. 5753203el Cytokine That Binds CD30
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kathryn A. Seese, Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: Apple 7.1
SOFTWARE: Microsoft Word, Version 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/580,014
FILING DATE: 20-DEC-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/225,989
FILING DATE: 12 APRIL 1994
APPLICATION NUMBER: US 07/966,775
FILING DATE: 27-OCT-1992
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 907,224

FILING DATE: 01-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 899,660
FILING DATE: 15-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 892,459
FILING DATE: 02-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 889,717
FILING DATE: 26-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Seese, Kathryn A.
REGISTRATION NUMBER: 32,172
REFERENCE/DOCKET NUMBER: 2804-E
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
TELEFAX: (206)233-0644
TELEX: 756822
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 125 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein fragment
HYPOTHEICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: hucd30 fragment (PRELIM)
US-08-580-014-20

Query Match 81.8%; Score 27; DB 1; Length 125;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKKXW 5
||| |
Db 81 FKKSW 85

RESULT 15
US-09-079-785-20
Sequence 20, Application US/09079785
Patent No. 6143869
GENERAL INFORMATION:
APPLICANT: Goodwin, Raymond G.
APPLICANT: Smith, Craig A.
APPLICANT: Arnltage, Richard J.
APPLICANT: Gruss, Hans-Jurgen
TITLE OF INVENTION: NO. 6143869el Cytokine That Binds CD30
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kathryn A. Seese, Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple 7.1
SOFTWARE: Microsoft Word, Version 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/079,785
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/225,989
FILING DATE: 12 APRIL 1994
APPLICATION NUMBER: US 07/966,775
FILING DATE: 27-OCT-1992

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 907,224
FILING DATE: 01-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 899,660
FILING DATE: 15-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 892,459
FILING DATE: 02-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 889,717
FILING DATE: 26-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: Seese, Kathryn A.
REGISTRATION NUMBER: 32,172
REFERENCE/DOCKET NUMBER: 2804-E
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
TELEFAX: (206)233-0644
TELEX: 756822
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 125 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein fragment
HYPOTHEICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: hucd30 fragment (PRELIM)
US-09-079-785-20

Query Match 81.8%; Score 27; DB 4; Length 125;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKKXW 5
||| |
Db 81 FKKSW 85

Search completed: February 27, 2002, 11:50:37
Job time: 866 sec.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:51:50 : Search time 66.26 Seconds
(without alignments)
5.590 Million cell updates/sec

Title: US-09-446-109A-22
Perfect score: 30
Sequence: 1 FPXWR 5

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_1101:*

- 1: /SID58/gcgdata/geneseq/geneseq/AA1980.DAT:*
- 2: /SID58/gcgdata/geneseq/geneseq/AA1981.DAT:*
- 3: /SID58/gcgdata/geneseq/geneseq/AA1982.DAT:*
- 4: /SID58/gcgdata/geneseq/geneseq/AA1983.DAT:*
- 5: /SID58/gcgdata/geneseq/geneseq/AA1984.DAT:*
- 6: /SID58/gcgdata/geneseq/geneseq/AA1985.DAT:*
- 7: /SID58/gcgdata/geneseq/geneseq/AA1986.DAT:*
- 8: /SID58/gcgdata/geneseq/geneseq/AA1987.DAT:*
- 9: /SID58/gcgdata/geneseq/geneseq/AA1988.DAT:*
- 10: /SID58/gcgdata/geneseq/geneseq/AA1989.DAT:*
- 11: /SID58/gcgdata/geneseq/geneseq/AA1990.DAT:*
- 12: /SID58/gcgdata/geneseq/geneseq/AA1991.DAT:*
- 13: /SID58/gcgdata/geneseq/geneseq/AA1992.DAT:*
- 14: /SID58/gcgdata/geneseq/geneseq/AA1993.DAT:*
- 15: /SID58/gcgdata/geneseq/geneseq/AA1994.DAT:*
- 16: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT:*
- 17: /SID58/gcgdata/geneseq/geneseq/AA1996.DAT:*
- 18: /SID58/gcgdata/geneseq/geneseq/AA1997.DAT:*
- 19: /SID58/gcgdata/geneseq/geneseq/AA1998.DAT:*
- 20: /SID58/gcgdata/geneseq/geneseq/AA1999.DAT:*
- 21: /SID58/gcgdata/geneseq/geneseq/AA2000.DAT:*
- 22: /SID58/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	96.7	10	AAV08336	O. hannah venom al
2	29	96.7	11	AAV08334	O. hannah venom al
3	29	96.7	38	AAV08491	Blinding domain of
4	29	96.7	94	AAV06599	Rat sodium channel
5	29	96.7	94	AAV06599	Rat truncated sodi
6	29	96.7	135	AAV04918	Mycobacterium spec
7	29	96.7	177	AAV04918	Breast and ovarian
8	29	96.7	299	AAV58892	Helicobacter pylori
9	29	96.7	314	AAV52488	CD46 construct del
10	29	96.7	324	AAV52488	CD46 from clone pm
11	29	96.7	359	AAV70396	Cellulase. Cellul

12	29	96.7	372	13	AAV27793	New platelet lacto
13	29	96.7	372	16	AAV92239	Chemokine superfam
14	29	96.7	372	16	AAV68813	Human lymphocyte p
15	29	96.7	372	20	AAV06644	Human BTK-like s ly
16	29	96.7	372	21	AAV90627	Human G protein-co
17	29	96.7	372	21	AAV90627	Human mutant G pro
18	29	96.7	373	12	AAV15230	CD46 from clone pm
19	29	96.7	376	17	AAV93939	CD46 wild-type. H
20	29	96.7	377	12	AAV15229	CD46 from clone pm
21	29	96.7	377	17	AAV93942	CD46 construct sub
22	29	96.7	377	17	AAV93941	Human MCP. Homo s
23	29	96.7	377	18	AAV27484	Human MCP. Homo s
24	29	96.7	384	12	AAV10924	Human CD46. Homo
25	29	96.7	418	21	AAV13589	Streptomyces globi
26	29	96.7	421	21	AAV58394	Lung cancer associ
27	29	96.7	421	22	AAV75528	Human colon cancer
28	29	96.7	497	20	AAV04972	C glutamicin spec
29	29	96.7	520	22	AAV90739	Porphyromonas ging
30	29	96.7	589	20	AAV34524	Human secreted pro
31	29	96.7	590	21	AAV94928	Porphyromonas ging
32	29	96.7	593	20	AAV34523	Porphyromonas ging
33	29	96.7	606	20	AAV34393	CAB-2 chimeric pro
34	29	96.7	611	22	AAV03762	Human protein sequ
35	29	96.7	612	22	AAV94498	H. pylori inner me
36	29	96.7	614	18	AAV20991	Human protein sequ
37	29	96.7	621	22	AAV94049	Human polypeptide
38	29	96.7	634	22	AAV40125	Human protein sequ
39	29	96.7	634	22	AAV92742	Human protein sequ
40	29	96.7	634	22	AAV83478	Human protein sequ
41	29	96.7	634	22	AAV95120	Human protein sequ
42	29	96.7	634	22	AAV95120	Human protein sequ
43	29	96.7	1464	19	AAV79294	An antigen from de
44	29	96.7	42	22	AAV15108	Peptide #1542 enco
45	28	93.3	42	22	AAV27563	Peptide #1600 enco

ALIGNMENTS

RESULT	1	
ID	AAV08336	standard: Peptide: 10 AA.
XX	AAV08336:	
AC	AAV08336:	
XX	19-JUL-1999 (first entry)	
DT	XX	
DE	XX	O. hannah venom alpha-neurotoxin homologous peptide 5.
KW	Snake: venom. King cobra: alpha-neurotoxin; toxin; analgesia; screening.	
XX	XX	Ophiophagus hannah.
OS	XX	
PN	XX	WO9924055-A1.
XX	XX	20-MAY-1999.
PD	XX	
XX	XX	03-NOV-1998: 98WO-SC00087.
PF	XX	
XX	XX	06-NOV-1997: 97SG-0003972.
PR	XX	
XX	XX	(UNST-) UNIT SINGAPORE NAT.
PA	XX	Gopaladrishtakone P, Gwee MCE, Kini RM, Pu XC, Wong PT;
XX	XX	WPI: 1999-327205/27.
DR	XX	
XX	XX	Snake neurotoxin derived peptides
XX	XX	Claim 11: Page 43: 46pp: English.
XX	XX	This invention describes novel peptide derivatives of alpha-neurotoxin,
XX	XX	derived from snake venom toxins of Ophiophagus hannah and which are

CC capable of inducing or facilitating analgesia. The peptides, optionally
 CC with a carrier and/or diluent, are used to induce analgesia in a subject.
 CC Antibodies derived from the products of the invention can be used to
 CC screen for subject peptides, e.g. for purification purposes.

XX Sequence 10 AA:

Query Match 96.7%: Score 29; DB 20; Length 10;
 Best Local Similarity 80.0%: Pred. No. 13;
 Matches 4: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPKWR 5
 1 1 1
 Db 3 fptwr 7

RESULT 2

AAV08334 standard; Peptide: 11 AA.

XX AAV08334:

DT 19-JUL-1999 (first entry)

DE O. hannah venom alpha-neurotoxin homologous peptide 3.

KW Snake: venom; king cobra; alpha-neurotoxin; toxin; analgesia; screening.

OS Ophiophagus hannah.

PN MO9924055-A1.

PD 20-MAY-1999.

PE 03-NOV-1998; 98WO-SC00087.

PR 06-NOV-1997; 97SG-0003972.

PA (UYST-) UNIV SINGAPORE MAT.

PI Gopaladrishnakone P., Gwee MCE, Kini RM, Pu XC, Wong PT;

DR WPI: 1999-327205/27.

XX Snake neurotoxin derived peptides

PS Claim 9; Page 42; 46pp: English.

CC This invention describes novel peptide derivatives of alpha-neurotoxin,
 CC derived from snake venom toxins of Ophiophagus hannah and which are
 CC capable of inducing or facilitating analgesia. The peptides, optionally
 CC with a carrier and/or diluent, are used to induce analgesia in a subject.
 CC Antibodies derived from the products of the invention can be used to
 CC screen for subject peptides, e.g. for purification purposes.

XX Sequence 11 AA:

Query Match 96.7%: Score 29; DB 20; Length 11;
 Best Local Similarity 80.0%: Pred. No. 14;
 Matches 4: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPKWR 5
 1 1 1
 Db 3 fptwr 7

RESULT 3

AAV08491 standard; Protein: 38 AA.

XX AAV08491:

DT 18-APR-1995 (first entry)

DE Binding domain of a polystyrene-binding TSAR (SB-9-5).

PN MO9418318-A.

PD 18-AUG-1994.

PE 01-FEB-1994; 94WO-US00977.

PR 01-FEB-1993; 93US-0013416.

PR 30-DEC-1993; 93US-0176500.

PR 31-JAN-1994; 94US-0189331.

PA (UYNC-) UNIV NORTH CAROLINA.

PI Fowlkes DM, Kay BK;

DR WPI: 1994-279739/34.

XX Identifying proteins or peptide(s) which bind a ligand - by
 screening a recombinant vector library expressing fusion proteins
 comprising a binding domain and an effector domain

PS Example 7.8; Page 120; 255pp: English.

XX AAV08491-507 are amino acid sequences of the binding domain of TSAR
 CC (Totally Synthetic Affinity Reagents) peptides. These particular
 CC peptides are encoded by the polystyrene (SB) binding phage. There
 CC is no apparent linear motif, but the peptides are rich in Trp, Tyr and
 CC Gly, poor in Arg, Val and Lys, and completely lack cysteine residues.
 CC The non-variable amino acids at the NH2 and COOH terminals are
 CC not shown. TSAR peptides are generated using generic oligonucleotides
 CC (see AAQ07470-73 for examples). TSARs are concatenated heterofunctional
 CC proteins or peptides, comprising at least two functional regions - a
 CC binding domain with affinity for a ligand and a second effector peptide
 CC portion that is chemically or biologically active. They may further
 CC comprise a linker peptide between the 2 domains. The TSARs or compns.
 CC comprising a TSAR binding domain can be used in vivo to deliver a
 CC chemically or biologically active moiety, eg. metal ion, radioisotope,
 CC peptide, toxin or enzyme, to the specific target or on the cell. They
 CC can also replace the function of macromolecules eg. monoclonal or
 CC polyclonal antibodies and therefore circumvent the need for complex
 CC methods of hybridoma formation or in vivo antibody production. The TSARs
 CC are easily characterised and have designed activity allowing direct and
 CC rapid detection in a screening process.

XX Sequence 38 AA:

Query Match 96.7%: Score 29; DB 15; Length 38;
 Best Local Similarity 80.0%: Pred. No. 46;
 Matches 4: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPKWR 5
 1 1 1
 Db 10 fptwr 14

RESULT 4

AAV06599 standard; Protein: 94 AA.

XX AAV06599:

```

XX 26-OCT-1999 (first entry)
XX
XX Rat sodium channel NaN splice variant C-terminal sequence.
DE
XX NaN; sodium channel; ion transport; rat; dorsal root ganglia;
XX pain; paraesthesia; hyperexcitability; therapy.
XX
XX Rattus sp.
XX
XX MO9938889-A2.
XX
XX 05-AUG-1999.
XX
XX 29-JAN-1999; 99WO-US02008.
XX
XX 20-NOV-1998; 98US-0109402.
XX
XX 29-JAN-1998; 98US-0072990.
XX
XX (UYVA ) UNIV YALE.
XX
XX Dlb-Hajj S, Waxman S;
XX
XX WPI: 1999-479168/40.
XX
XX
XX New isolated nucleic acids encoding sodium channels, used to develop
XX products for treating acute or chronic pain or hyperexcitability
XX phenomena
XX
XX Example 5: Page 35; 91pp; English.
XX
XX This is the C-terminal sequence of a splice variant of rat sodium
XX channel NaN. This truncated variant comprises amino acids 1-1378
XX of the NaN sequence given in AY06596 and the present 94-amino acid
XX sequence at the C-terminal end. It thus lacks the 387 C-terminal
XX amino acids of full-length NaN. The variant arises from the use
XX of a cryptic donor splice site in exon 23 and a novel exon 23'
XX which is located in intron 23 of the rat NaN gene. NaN is a
XX previously unidentified voltage gated sodium channel protein that
XX is preferentially expressed in dorsal root ganglia or trigeminal
XX ganglia, and which produces a TRX-R sodium current. The invention
XX provides methods for identifying agents that modulate NaN channel
XX activity or expression and for using such agents to treat acute or
XX chronic pain, paraesthesia and hyperexcitability phenomena.
XX
XX Sequence 94 AA:
SQ

```

Query Match 96.7%; Score 29; DB 20; Length 94;
 Best Local Similarity 80.0%; Pred. No. 1.1e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

OY 1 FPXWR 5
DB 52 fpawr 56

```

RESULT 5
 AAB20130
 ID AAB20130 standard; Protein: 94 AA.
 XX
 AC AAB20130;
 XX
 DT 30-APR-2001 (first entry)
 XX
 DE Rat truncated sodium channel C-terminal region.
 XX
 KW Sodium channel; NaN; rat; tetrodotoxin resistant; pain;
 KW paraesthesia; hyperexcitability; analgesic; vaccine; therapy;
 KW diagnosis.
 XX
 OS Rattus norvegicus.
 XX

```

PN MO200105831-A1.
XX
XX 25-JAN-2001.
XX
XX 14-JUL-2000; 2000WO-US19342.
XX
XX 16-JUL-1999; 99US-0354147.
XX
XX (UYVA ) UNIV YALE.
XX
XX Dlb-Hajj S, Waxman SG;
XX
XX WPI: 2001-103147/11.
XX
XX
XX Nucleic acid molecules encoding human tetrodotoxin resistant sodium
XX channels, useful for preventing, diagnosing and treating pain,
XX paraesthesia and/or hyperexcitability phenomena -
XX
XX Example 5: Page 34; 162pp; English.
XX
XX The present sequence is that of the C-terminal portion of a
XX truncated rat NaN variant. The variant lacks the 387 C-terminal
XX amino acids of full-length rat NaN (see AAB20122), which are
XX replaced by the present 94-amino acid polypeptide. The N-terminal
XX 1378 amino acids of the truncated variant are identical to those
XX of full-length NaN. The new sequence arises from the use of a
XX cryptic donor splice site in exon 23 and a novel exon 23' which is
XX located in intron 23. Rat NaN is a novel tetrodotoxin resistant
XX sodium channel NaN belonging to the a-subunit voltage-gated sodium
XX channel protein family. These sodium channels underlie the
XX generation and propagation of impulses in excitable cells such as
XX neurons and muscle fibres. Preferential expression of NaN on
XX sensory neurons innervating the body (dorsal root ganglia) and
XX the face (trigeminal ganglia), but not on other neurons, makes
XX it a very useful target for diagnostic and/or therapeutic uses in
XX relation to acute and/or chronic pain pathologies. A claimed
XX method of treating pain, paraesthesia and/or hyperexcitability
XX phenomena in a human or animal subject involves administering an
XX agent that alters sodium current flow through NaN channels, or
XX which modulates transcription or translation of NaN mRNA, in
XX dorsal root ganglia or trigeminal neurons. NaN polypeptides can
XX be obtained by recombinant expression, and used to treat disorders
XX associated with decreased sodium channel expression, to screen for
XX compounds that modulate sodium channel expression or activity,
XX and to raise antibodies useful as diagnostic agents.
XX
XX Sequence 94 AA:
SQ

```

Query Match 96.7%; Score 29; DB 22; Length 94;
 Best Local Similarity 80.0%; Pred. No. 1.1e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

OY 1 FPXWR 5
DB 52 fpawr 56

```

RESULT 6
 AAY04918
 ID AAY04918 standard; Protein: 135 AA.
 XX
 AC AAY04918;
 XX
 DT 06-JUL-1999 (first entry)
 XX
 DE Mycobacterium species protein sequence 35B.
 XX
 KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
 KW hybridisation; detection; vaccine; immunisation; infection.
 XX
 OS Mycobacterium sp.
 XX

PN MO9909186-A2.
 XX 25-FEB-1999.
 PD
 XX
 PF 14-AUG-1998; 98MO-FR01813.
 XX
 PR 11-SEP-1997; 97FR-0011325.
 XX 14-AUG-1997; 97FR-0010404.
 XX
 PA (INSP) INST PASTEUR.
 XX
 PI Gicquel, B., Lim EM, Pellicic V, Portnoi D, Goguet de la Salmoniere Y;
 PI Guigueno, A.;
 XX
 DR WPI: 1999-181045/15.
 N-PSDB: AAX34171.
 XX
 CC Mycobacterial DNA vectors containing reporter constructs - for
 PT identifying coding or promoter sequences involved in
 PT infection-associated protein expression
 CC
 PS Claim 32; Fig 35B; 309pp; French.
 XX
 CC Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted
 CC proteins from various Mycobacterium species microorganisms. The
 CC encoding nucleotide sequences can be used as primers and probes for
 CC methods for detecting and identifying mycobacteria, especially belonging
 CC to the M. tuberculosis complex. The encoded proteins can be used in
 CC vaccines for immunisation against a bacterial or viral infection.
 CC
 SQ Sequence 135 AA;
 XX
 QY 1 FPXMR 5
 1111
 Db 66 fpawr 70
 XX
 RESULT 7
 AAB58892
 ID AAB58892 standard; Protein: 177 AA.
 XX
 AC AAB58892;
 XX
 DT 27-MAR-2001 (first entry)
 XX
 DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 600.
 XX
 KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
 KW neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 KW antidiabetic; antineoplastic; antitumor; antitubercular; anticonvulsant;
 KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
 KW Addison's disease; allergy; autoimmune haemolytic anaemia;
 KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
 KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
 KW cardiovascular disorder; wound healing; neurological disease.
 KW
 OS Homo sapiens.
 XX
 XX WO20005173-A1.
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05881.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX

PI Rosen CA, Ruben SM;
 XX
 DR WPI: 2000-611515/58.
 DR N-PSDB: AAF21795.
 XX
 XX New human breast and ovarian cancer associated gene sequences and the
 PT polypeptides encoded by these genes, useful in the prevention,
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular
 PT disorders and neurological diseases -
 XX
 PS Claim 11; Page 1037-1038; 1299pp; English.
 XX
 CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
 CC associated with breast and ovarian cancer. Included in the invention are
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
 CC isolation and characterisation of the DNA and protein sequences of the
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist
 CC or antagonist sequences exhibit cytostatic; immunosuppressive;
 CC neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 CC antidiabetic; antineoplastic; antitumor; antitubercular; anticonvulsant;
 CC antibacterial; antifungal; antiparasitic and cardiant activity. The
 CC polynucleotide and protein sequences are used in the diagnosis of cancer,
 CC particularly breast and ovarian cancer. The nucleic acid sequences,
 CC proteins, agonists and antagonists may also be used in the diagnosis,
 CC prevention and treatment of immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; cardiovascular disorders such as
 CC myocardial ischaemia; wound healing; neurological diseases such as
 CC cerebral anoxia and epilepsy; and infectious diseases.
 CC
 SQ Sequence 177 AA;
 XX
 QY 1 FPXMR 5
 1111
 Db 22 fpawr 26
 XX
 RESULT 8
 AAB52488
 ID AAB52488 standard; Protein: 299 AA.
 XX
 AC AAB52488;
 XX
 DT 23-FEB-2001 (first entry)
 XX
 DE Helicobacter pylori bait polypeptide #6.
 XX
 KW Helicobacter pylori; two-hybrid system; protein-protein interaction;
 KW bait polypeptide; gastric ulcer; antibacterial.
 KW
 OS Helicobacter pylori.
 XX
 PN WO200006722-A1.
 XX
 PD 09-NOV-2000.
 XX
 PF 14-APR-2000; 2000WO-1B00603.
 XX
 PR 30-APR-1999; 99EP-0401066.
 XX
 PA (HYBR-) HYBRIGENICS SA.
 XX
 XX Legrain P, Selig L, Rain J;
 PI WPI: 2000-687535/67.
 DR N-PSDB: AAC97234.
 DR

XX A two-hybrid system for identifying compounds useful in the treatment
 PT of e.g. gastric ulcers comprises producing a collection of recombinant
 PT cell clones -
 XX
 PS Example 5: Page 88-89; 267pp: English.
 CC The present sequence is a bait polypeptide used in a Helicobacter
 CC pylori two-hybrid screen to identify protein-protein interactions.
 CC The method is used to identify a recombinant cell clone expressing a
 CC prey polypeptide which is capable of interacting with the bait
 CC polypeptide. The two hybrid system is useful for screening compounds
 CC for antibacterial activity. It may be used in the treatment of gastric
 CC ulcers. The polynucleotides are useful as amplification primers or
 CC specific detection probes. The polypeptides, vectors or host cells can
 CC be used as immunogens to produce mono- or polyclonal antibodies. The
 CC polynucleotides, polypeptides, antibodies, vectors, host cells or
 CC modulating agents can be used to produce a pharmaceutical composition.
 XX
 SQ Sequence 299 AA:

Query Match 96.7%; Score 29; DB 21; Length 299;
 Best Local Similarity 80.0%; Pred. No. 3.3e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
 1111
 Db 89 fpswr 93

RESULT 9

AAR93943
 ID AAR93943 standard; Protein; 314 AA.

XX AAR93943:

DT 21-MAY-1996 (first entry)

DE CD46 construct delSCR2/subSCR3.

XX CD46; recombinant protein; short consensus repeat; SCR:
 KW regulator of complement activation; transgenic animal; pig;
 KM organ transplantation.
 XX Synthetic.

OS Synthetic.

PN WO9606937-A1.

PD 07-MAR-1996.

PF 30-AUG-1995; 95MO-AU00553.

PR 30-AUG-1994; 94AU-0007724.

PA (AUST-) AUSTIN RES INST.

PI Christiansen D, Loveland B, McKenzie JFC, Milland J;

WPI: 1996-160368/16.

DR N-PSDB; AAT17599.

PT Increasing prodn. of recombinant proteins, esp. CD46 - by reducing
 PT the amt. of A and/or T in an A and/or T rich region of encoding gene
 PT exon
 XX

PS Claim 12: Page 40-41; 60pp: English.

CC CD46 delSCR2/subSCR3 (AAR93943) is the product of a cDNA construct
 CC (AAT17599) obtd. by splice overlap extension PCR of wild-type CD46
 CC cDNA (AAT17595). The A-T content of A-T-rich exon 5 of the gene,
 CC encoding the short consensus repeat (SCR) 3 of CD46, was lowered
 CC and SCR2 was deleted. This resulted in improved prodn. in

CC eukaryotic host cells, e.g. CHO-K1 and COS-7. CD46 delSCR2/subSCR3
 CC is used to prevent complement- or inflammation-mediated tissue
 CC damage, to improve immunity to tumours or viruses, to control
 CC fertilisation and to prevent spontaneous abortion. Expression in
 CC transgenic animals, esp. pigs, provides organs suitable for
 CC transplantation.
 XX
 SQ Sequence 314 AA:

Query Match 96.7%; Score 29; DB 17; Length 314;
 Best Local Similarity 80.0%; Pred. No. 3.4e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
 1111
 Db 11 fpswr 15

RESULT 10

AAR15233
 ID AAR15233 standard; Protein; 324 AA.

XX AAR15233:

DT 17-MAR-1992 (first entry)

DE CD46 from clone pm5.8.

XX SCR: short consensus repeats; transmembrane; cytoplasmic;
 KW membrane cofactor protein; MCP.
 XX Homo sapiens.

XX Homo sapiens.

PH Key Location/Qualifiers

FT Peptide 1..34

FT Protein 34..324

FT Modified-site 83

FT Modified-site 114

FT Domain 289..304

FT /label= N-glycosylation_site

FT /label= N-glycosylation_site

FT /label= hydrophobic-transmembrane_domain

PN WO9118097-A.

PD 28-NOV-1991.

PF 10-MAY-1991; 91MO-AU00199.

PR 11-MAY-1990; 90AU-0000133.

PA (UYME-) UNIV MELBOURNE.

PI Purcell DFJ, Russell SM, McKenzie JFC;

WPI: 1991-369251/50.

DR N-PSDB; AAQ14919.

PT New CD46 membrane co-factor protein variants - useful as probes
 PT to identify CD46 isoforms and for diagnosing spontaneous
 PT abortion, inhibiting immuno-response and treating leukemia
 XX

PS Disclosure: Fig 1A and 3B: 77pp: English.

CC The sequence of the pm5.8 clone is identical to the pm5.1 clone in
 CC the portion encoding the NH2 leader and four SCR regions. The
 CC sequence after nucleotide 890 was different, and results from
 CC reading through of the cDNA into an intron sequence after the
 CC fourth SCR. The protein encoded by this new sequence encodes
 CC a 16 amino acid hydrophobic region, probably serving as a membrane

CC spanning sequence. The putative cytoplasmic tail would also
 CC contain a high proportion of charged amino acids.
 CC See also AAQ14915-25, AAQ15211-12 and AAR15457-59.
 XX
 SO Sequence 324 AA;

Query Match 96.7%; Score 29; DB 12; Length 324;
 Best Local Similarity 80.0%; Pred. No. 3.5e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 11 f p s w r 15

RESULT 11

AAp70396 standard; Protein: 359 AA.

AC AAp70396;

DT 30-APR-1991 (first entry)

DE Cellulase.

KW Cellulase; plasmid; transformation.

OS Cellulomonas uda C04 (FER.

XX JP62175178-A.

PD 31-JUL-1987.

PE 30-JAN-1986; 86JP-0016800.

PR 30-JAN-1986; 86JP-0016800.

PA (ENER-) SHIN ENERGY SOGO KAIHATSU KIKO.

PA (NENR-) NENRYOYO ALCOHOL KAIHATSU GIJUTSU KENKYU KUMIAT.

XX WPI: 1987-253827/36.

DR N-PSDB: AAN70617.

PT Bio-engineering cellulase prodn. - by preparing DNA chain,
 PT preparing plasmid contg. DNA chain in the gene expressible state,
 PT and transforming bacterial by plasmid

PS Disclosure: Fig 1(A-B): 12pp; Japanese.

CC Transformation E.coli JM103 (FERM P-8612) with the gene encoding
 CC this protein allows expression of the protein and easy and cheap
 CC prodn. of cellulase.

XX Sequence 359 AA;

SO

Query Match 96.7%; Score 29; DB 8; Length 359;
 Best Local Similarity 80.0%; Pred. No. 3.9e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 11 f p s w r 15

RESULT 12

AAr27793 standard; Protein: 372 AA.

AC AAr27793;

DT 12-MAR-1993 (first entry)

XX New platelet factor 4 receptor superfamily member PF4AR1.
 DE
 XX IL-8R; G-protein coupled receptor family; rhodopsin superfamily;
 KM pro-inflammatory cytokine; 8rr.9.
 KW
 XX

OS Homo sapiens.
 XX
 PN W09217497-A.
 XX
 PD 15-OCT-1992.

XX 23-MAR-1992; 92MO-US02317.

FE 29-MAR-1991; 91US-0677211.

FR 19-DEC-1991; 91US-0810782.

XX (GETH) GENENTECH INC.

XX Holmes WE, Lee J, Wood WI;

DR WPI: 1992-366191/44.

JR N-PSDB: AAQ37107.

XX Isolated human platelet factor 4 super-family receptor
 PT polypeptide and corresp. antibodies and DNA - useful as
 PT diagnostic and screening agents, and for treating inflammation or
 PT PF4AR-mediated disorders

XX Claim 7; Fig 5; 78pp; English.

XX The IL-8 receptor cDNA sequence was isolated (see AAQ29505) and a
 CC 874bp sub-fragment of the coding sequence was used as a probe to
 CC screen human cell line HL60 and human peripheral blood lymphocyte
 CC cDNA libraries. Two new gene sequences were found that are clearly
 CC related to the IL-8 receptor. One of these was contained in clone
 CC 8rr.9 and is predicted to encode an amino acid sequence which is
 CC 36% and 38% identical with the high and low affinity IL-8 receptor
 CC sequences, respectively. See also AAQ37107.

XX Sequence 372 AA;

SO

Query Match 96.7%; Score 29; DB 13; Length 372;
 Best Local Similarity 80.0%; Pred. No. 4e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 11 f p s w r 15

RESULT 13

AAr92239 standard; Protein: 372 AA.

AC AAr92239;

DT 26-MAR-1996 (first entry)

DE Chemokine superfamily receptor.

XX Interleukin; IL-8; inflammation; psoriasis; dermatitis;

KW rheumatoid arthritis; inflammatory bowel disease;

FW chronic lung inflammation; treatment; antibody;

XX affinity purification; detection.

XX Homo sapiens.

US US5440021-A.
 XX 08-AUG-1995.
 XX

PF 29-MAR-1991; 91US-0677211.
 PR 25-FEB-1994; 94US-0202056.
 PR 29-MAR-1991; 91US-0677211.
 XX
 PA (CHUN/J) CHUNTHARAPAI A.
 PA (HEBE/J) HEBERT C.
 PA (KIM/J) KIM K J.
 PA (LEE/J) LEE J.
 XX
 PI Chuntharapai A, Hebert C, Kim KJ, Lee J;
 XX
 DR WPI: 1995-283151/37.
 DR N-PSDB: AAQ99009.
 XX
 PT New antibodies against interleukin 8 type B receptor - used to treat
 PT or prevent inflammation, also for detecting receptor expression and
 PT purification.
 XX
 PS Example 2; Columns 49-52; 62pp; English.
 XX
 CC Antibodies directed against the interleukin-8 receptor B can be used
 CC to treat or prevent inflammation e.g. psoriasis, dermatitis,
 CC rheumatoid arthritis and particularly inflammatory bowel disease and
 CC chronic lung inflammation. When immobilised, these antibodies may
 CC be used to detect interleukin-8 receptor B expression in cells and
 CC tissues and for affinity purification of interleukin-8 receptor B
 CC from cells. This sequence is an additional chemokine superfamily
 CC receptor which was identified by probing lambda libraries of genomic
 CC DNA from a human monocyte-like cell line (U-60) and human peripheral
 CC blood lymphocytes using a large fragment of the interleukin-8 type
 CC A receptor DNA (See AAQ99006).
 CC
 SQ Sequence 372 AA:

Query Match 96.7%; Score 29; DB 16; Length 372;
 Best Local Similarity 80.0%; Pred. No. 4e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 1 FPXMR 5
 1111
 Db 352 fpssvr 356

RESULT 14

ID AAR68813 standard; Protein: 372 AA.
 XX

AC AAR68813;

DT 18-JUL-1995 (first entry)

DE Human lymphocyte PF4AR.

KM Interleukin-8 receptor; IL-8 receptor; PF4AR;

KM platelet factor superfamily receptor; lymphocyte; chemotactic;
 inflammation; inflammatory disease; arthritis; emphysema; cystic;
 fibrosis; colitis; bronchitis; meningitis; therapeutic.

OS Homo sapiens.

PN WO9428931-A.

PD 22-DEC-1994.

PF 07-JUN-1994; 94MO-US06380.

PR 11-JUN-1993; 93US-0076093.

PA (GETI) GENENTECH INC.

PI Chuntharapai A, Hebert C, Kim KJ, Lee J;

XX WPI: 1995-036114/05.
 DR N-PSDB: AA080522.
 XX
 PT Treatment of inflammatory disorders - by administering an
 PT antibody capable of binding a platelet factor 4 superfamily
 PT receptor polypeptide
 XX
 PS Disclosure: Page 56-58; 83pp; English.
 XX
 CC 2 PF4AR members were identified by probing lambda libraries from
 CC human monocyte-like cell line HL-60 and human peripheral blood
 CC lymphocytes using a large fragment of IL-8 receptor DNA (full
 CC sequence given in AA080520). The nucleotide sequences of the 2
 CC PF4ARs are given in AA080521 and AA080522, and their respective
 CC amino acid sequences in AAR68812 and AAR68813.
 CC
 SQ Sequence 372 AA:

Query Match 96.7%; Score 29; DB 16; Length 372;
 Best Local Similarity 80.0%; Pred. No. 4e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 1 FPXMR 5
 1111
 Db 352 fpssvr 356

RESULT 15

ID AAY06644 standard; Protein: 372 AA.
 XX

AC AAY06644;

DT 26-OCT-1999 (first entry)

DE Human Burkitt's lymphoma receptor 1 (BLR1).

KM Burkitt's lymphoma receptor 1; BLR1; human;

KM B lymphocyte chemottractant; BLC; chemokine; ligand;
 drug screening; leukaemia; autoimmune disease; therapy.

OS Homo sapiens.

PN WO9928468-A1.

PD 10-JUN-1999.

PF 02-DEC-1998; 98MO-US25561.

PR 02-DEC-1997; 97US-0982493.

PA (REGC) UNIV CALIFORNIA.

PI Cyster JC, Gunn MD, Williams LT;

DR WPI: 1999-493764/41.

DR N-PSDB: AAX87710.

PT Modulating interaction of a Burkitt's lymphoma Receptor 1
 PT polypeptide and ligand, useful in drug screens

PS Claim 1; Page 38; 42pp; English.

This sequence represents human Burkitt's lymphoma receptor 1
 (BLR1). The invention relates to methods for modulating the
 interaction of BLR1 with its ligand, B lymphocyte chemottractant
 (BLC, see AAY06642). The methods comprise combining BLR1 and BLC
 polypeptides with a candidate modulator agent under conditions
 whereby, but for the presence of the agent, the polypeptides engage
 in a first interaction, and determining a second interaction of the
 polypeptides in the presence of the agent, wherein a difference

CC between the first and second interactions indicates that the agent
CC modulates the interaction of the polypeptides. The modulator is
CC preferably an antagonist, especially dominant negative, form of
CC BIC. BIC and BIC agonists and antagonists may be useful for
CC treating viral (e.g. HIV) infections, lymphoma, B lineage
CC leukaemia, and autoimmune diseases such as rheumatoid arthritis,
CC thyroiditis and diabetes.
XX
SQ Sequence 372 AA;

Query Match 96.7%; Score 29; DB 20; Length 372;
Best Local Similarity 80.0%; Pred. No. 4e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FPXWR 5
| | | |
Db 352 fpxwr 356

Search completed: February 27, 2002, 11:51:50
Job time: 627 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:52:58 ; Search time 34.86 Seconds

(without alignments)
10.926 Million cell updates/sec

Title: US-09-446-109a-22

Perfect score: 30

Sequence: 1 FPXMR 5

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.68:.*
2: PIR2:.*
3: PIR3:.*
4: PIR4:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	96.7	73	1 N20H1	long neurotoxin 1
2	29	96.7	113	2 T07855	translational elonga
3	29	96.7	157	2 T24380	hypothetical prote
4	29	96.7	245	2 D82680	conserved hypothet
5	29	96.7	256	2 D85510	probable EC 3.5. a
6	29	96.7	259	2 A29831	heat-labile entero
7	29	96.7	273	2 T05454	hypothetical prote
8	29	96.7	292	2 C82960	hypothetical prote
9	29	96.7	327	2 S56162	MCRI5 protein - h
10	29	96.7	332	2 T05855	minor endoglucanas
11	29	96.7	348	2 C02913	sperm CD46 - human
12	29	96.7	359	2 I40696	endoglucanase - Ce
13	29	96.7	369	2 I57998	membrane cofactor
14	29	96.7	372	2 S26667	C protein-coupled
15	29	96.7	377	2 I54479	membrane cofactor
16	29	96.7	384	2 S01896	membrane cofactor
17	29	96.7	417	2 D86251	protein F25C20.9 (
18	29	96.7	479	2 T35441	aldehyde dehydroge
19	29	96.7	497	2 D83628	probable aldehyde
20	29	96.7	555	2 F69312	heterodisulfide re
21	29	96.7	625	2 H70330	hypothetical prote
22	29	96.7	642	2 D71909	ferrous iron trans
23	29	96.7	642	2 C64605	iron(II) transport
24	29	96.7	630	2 T22002	hypothetical prote
25	29	96.7	735	2 A83006	hypothetical prote
26	29	96.7	1451	2 I40325	dermoecrotic toxin
27	28	93.3	68	2 S75058	transposase ssr289
28	28	93.3	84	2 S76091	hypothetical prote
29	28	93.3	84	2 S76443	hypothetical prote

30	28	93.3	93	2 S75008	transposase ss1192
31	28	93.3	103	2 T47718	hypothetical prote
32	28	93.3	114	2 S77061	transposase sl1066
33	28	93.3	143	1 HHK48	heat shock protein
34	28	93.3	143	1 HHK41	heat shock protein
35	28	93.3	147	2 A71560	hypothetical prote
36	28	93.3	149	2 T35846	probable integral
37	28	93.3	157	2 F81710	conserved hypothet
38	28	93.3	169	2 C69300	conserved hypothet
39	28	93.3	254	2 S76459	transposase sl1186
40	28	93.3	254	2 E82791	conserved hypothet
41	28	93.3	258	2 S45682	acetone-cyanhydrin
42	28	93.3	259	2 S76643	transposase sl1051
43	28	93.3	261	2 S77171	transposase sl1171
44	28	93.3	261	2 S77351	transposase sl1171
45	28	93.3	261	2 S75081	transposase sl1026

ALIGNMENTS

RESULT 1

N20H1

Long neurotoxin 1 - king cobra

N:Alternate names: neurotoxin A

C:Species: Ophiophagus hannah (king cobra)

C:Date: 22-Jun-1981 #sequence_revision 22-Jun-1981 #text_change 23-Aug-1996

C:Accession: A01658

R:Joubert, F.J.

Biochim. Biophys. Acta 317, 85-98, 1973

A:Title: Snake venom toxins. The amino acid sequences of two toxins from ophiophagus

A:Reference number: A90588; MUID:73231298

A:Accession: A01658

A:Molecule type: protein

A:Residues: 1-73 <JOU>

C:Superfamily: snake toxin

C:Keywords: neurotoxin; venom

F:3-21,14-42,27-31,46-57,58-63/Disulfide bonds: #status predicted

Query Match 96.7% Score 29; DB 1; Length 73;
Best Local Similarity 80.0%; Pred. No. 33;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5

Db 66 FPXMR 70

RESULT 2

T07855

translational elongation factor eEF-1 alpha chain - rape (fragment)

N:Alternate names: translational elongation factor EF-1A

C:Species: Brassica napus (rape)

C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 31-Mar-2000

C:Accession: T07855

R:Saez-Vasquez, J.

submitted to the EMBL Data Library, February 1995

A:Reference number: Z16177

A:Accession: T07855

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-113 <SAE>

A:Cross-references: EMBL:U21744; NID:9914912; PIDD:AAA86366.1; PID:9914913

C:Experimental source: cv. Samourai; 5 days old etiolated seedlings

C:Superfamily: translational elongation factor Tu; translational elongation factor Tu hom

C:Keywords: GTP binding; protein biosynthesis

Query Match 96.7% Score 29; DB 2; Length 113;
Best Local Similarity 80.0%; Pred. No. 50;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FPXWR 5
Db 103 FPMWR 107

RESULT 3

T24380
hypothetical protein T03D8.2 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24380
R:Mortimore, B.
submitted to the EMBL Data Library, March 1997
A:Reference number: Z19884
A:Accession: T24380
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-157 <MIL>
A:Cross-references: EMBL:Z293838; PDB:CAH07406.1; GSPDB:GN00023; CESP:T03D8.2
A:Experimental source: clone T03D8
C:Genetics:
A:Gene: CESP:T03D8.2
A:Map position: 5
A:Introns: 49/3; 83/1; 125/3

Query Match 96.7%; Score 29; DB 2; Length 157;
Best Local Similarity 80.0%; Pred. No. 68;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FPXWR 5
Db 40 FPMWR 44

RESULT 4

D82680
conserved hypothetical protein XF1438 [imported] - *Xylella fastidiosa* (strain 9a5c)
C:Species: *Xylella fastidiosa*
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: D82680
R:anonymous. The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequences
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
A:Reference number: A82515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: D82680
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-245 <STM>
A:Cross-references: GB:AE003974; GB:AE003849; NID:g9106454; PDB:AAE84247.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Slipson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Brites, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, I
as-Neto, E.; Docena, C.; El-Dorfy, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to Genbank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Klieger, J.E.; Kuramae, E.E.; Laig
chado, M.A.; Medeira, A.M.B.N.; Medeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, F
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmiter, D.
Rodrigues, V.; Rosa, A.J.; de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.;
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1438

Query Match 96.7%; Score 29; DB 2; Length 245;
Best Local Similarity 80.0%; Pred. No. 1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FPXWR 5
Db 197 FPMWR 201

RESULT 5

D85510
probable EC 3.5. amylase-type enzyme yafV [imported] - *Escherichia coli* (strain O157;
C:Species: *Escherichia coli*
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: D85510
R:Pena, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
hler, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoulis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: D85510
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-256 <STO>
A:Cross-references: GB:AE005174; NID:g12512987; PDB:AAG54544.1; GSPDB:GN00145; UMG:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: yafV

Query Match 96.7%; Score 29; DB 2; Length 256;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FPXWR 5
Db 244 FPMWR 248

RESULT 6

A29831
heat-labile enterotoxin IIA chain A precursor - *Escherichia coli*
R:Alternate names: LT-IIa
C:Species: *Escherichia coli*
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C:Accession: A29831
R:Pickett, C.L.; Weinstein, D.L.; Holmes, R.K.
J. Bacteriol. 169, 5180-5187, 1987
A:Title: Genetics of type IIA heat-labile enterotoxin of *Escherichia coli*: operon *fus*
A:Reference number: A91849; MUID:88032841
A:Accession: A29831
A:Molecule type: DNA
A:Residues: 1-259 <PIC>
A:Cross-references: GB:M17894; NID:g146671; PDB:AAA24093.1; PDB:g146672
A:Note: the authors translated the codon TAT for residue 225 as Thr
C:Superfamily: heat-labile enterotoxin chain A
E:118/Domain: signal sequence #status predicted <SIG>
E:19-259/Product: heat-labile enterotoxin IIA chain A #status predicted <LTA>

Query Match 96.7%; Score 29; DB 2; Length 259;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FPXWR 5
Db 187 FPMWR 191

RESULT 7

T05454
hypothetical protein F7K2.180 - *Arabidopsis thaliana*
C:Species: *Arabidopsis thaliana* (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Sep-1999
C:Accession: T05454
R:Bevan, M.; Wedler, H.; Wandut, R.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Schue

submitted to the Protein Sequence Database, November 1998
A:Reference number: Z15416

A:Accession: T05454
A:Molecule type: DNA

A:Residues: 1-273 <BEV>

A:Cross-references: EMBL:AL033545

A:Experimental source: cultivar Columbia; BAC clone F7K2

A:Genetics:

A:Map position: 4

A:Note: F7K2.180

C:Superfamily: Arabidopsis hypothetical protein F7K2.180

Query Match 96.7%: Score 29; DB 2; Length 273;
Best Local Similarity 80.0%: Pred. No. 1.1e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPMXR 5

Db 79 FPMXR 83

RESULT 8

C82960
hypothetical protein PA5488 [Imported] - Pseudomonas aeruginosa (strain PAO1)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: G82960

R:Stover, C.K.; Plam, X.O.; Ervin, A.L.; Micoquich, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

..; Loty, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho

A:Reference number: A82950; MUID:20437337

A:Accession: G82960

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-292 <STO>

A:Cross-references: GB:AE004961; GB:AE004091; NID:9951814; PIDN:AMC08873.1; GSPDB:GN001

A:Experimental source: strain PAO1

C:Genetics:

A:Gene: PA5488

Query Match 96.7%: Score 29; DB 2; Length 292;
Best Local Similarity 80.0%: Pred. No. 1.2e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPMXR 5

Db 231 FPMXR 235

RESULT 9

S56162
MDCK15 protein - human

C:Species: Homo sapiens (man)

C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999

C:Accession: S56162

R:Barrella, L.; Loetscher, M.; Tobler, A.; Baggiolini, M.; Moser, B.

Biochem. J. 309, 773-779, 1995

A:Title: Sequence variation of a novel heptahelical leucocyte receptor through alternat

A:Reference number: S56162; MUID:95366951

A:Accession: S56162

A>Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-327 <BAR>

A:Cross-references: EMBL:X68829; NID:9840783; PIDN:CAA48723.1; PTD:9840784

C:Superfamily: vertebrate rhodopsin

Query Match 96.7%: Score 29; DB 2; Length 327;
Best Local Similarity 80.0%: Pred. No. 1.4e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPMXR 5

Db 307 FPMXR 311

RESULT 10

JT0585

minor endoglucanase (EC 3.2.1.-) precursor - Erwinia chrysanthemi

C:Species: Erwinia chrysanthemi

C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 08-Oct-1999

C:Accession: JT0585; JT0586

R:Guiseppi, A.; Aymeric, J.L.; Caml, B.; Barras, F.; Creuzet, N.

Gene 106, 109-114, 1991

A:Title: Sequence analysis of the cellulase-encoding cely gene of Erwinia chrysanthem

A:Reference number: JT0585; MUID:92039050

A:Accession: JT0585

A:Molecule type: DNA

A:Residues: 1-332 <GUL>

A:Cross-references: GB:M74044; NID:q148391; PIDN:AAA24818.1; PID:q148392

A:Experimental source: strain 3937

A:Accession: JT0586

A:Molecule type: protein

A:Residues: 24-33 <GUL1>

C:Genetics:

A:Gene: cely

C:Keywords: glycosidase; hydrolase

F:1-23/Domain: signal sequence #status predicted <SIG>

F:24-33/Product: minor endoglucanase #status predicted <MIN>

Query Match 96.7%: Score 29; DB 2; Length 332;
Best Local Similarity 80.0%: Pred. No. 1.4e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPMXR 5

Db 178 FPMXR 182

RESULT 11

G02913
sperm CD46 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 09-Jun-2000

C:Accession: G02913

R:Hara, T.

submitted to Genbank, March 1996

A:Reference number: H01942

A:Accession: G02913

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-349 <HAR>

A:Cross-references: GB:D84105; NID:q1256700; PIDN:BA12224.1; PID:q1256701

C:Superfamily: herpesvirus complement control protein; complement factor H repeat hom

F:35-94/Domain: complement factor H repeat homology <FH01>

F:99-157/Domain: complement factor H repeat homology <FH02>

F:162-223/Domain: complement factor H repeat homology <FH03>

F:228-283/Domain: complement factor H repeat homology <FH04>

Query Match 96.7%: Score 29; DB 2; Length 349;
Best Local Similarity 80.0%: Pred. No. 1.4e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPMXR 5

Db 11 FPMXR 15

RESULT 12

I40696

endoglucanase - Cellulomonas uda
C:Species: Cellulomonas uda
C>Date: 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 05-Nov-1999
C:Accession: 140696
R:Nakamura, K.; Misawa, N.; Kitamura, K.
J. Biotechnol. 4: 247-254, 1986
A:Title: Sequence of a cellulase gene of Cellulomonas uda CB4.
A:Reference number: 140696
A:Accession: 140696
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-359 <RES>
A:Cross-references: GB:M36503; NID:g144430; PIDN:AAA23090.1; PID:g144431

Query Match 96.7%: Score 29; DB 2; Length 359;
Best Local Similarity 80.0%: Pred. No. 1.5e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
11 11
Db 178 FPMWR 182

RESULT 13
157998
membrane cofactor protein - human
C:Species: Homo sapiens (man)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 19-May-2000
C:Accession: 157998
R:Cervoni, F.; Fenchel, P.; Akhoundi, C.; Hsl, B.L.; Rossi, B.
Mol. Reprod. Dev. 34: 107-113, 1993
A:Title: Characterization of a cDNA clone coding for human testis membrane cofactor protein.
A:Reference number: 157998; MUID:93119658
A:Accession: 157998
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-369 <RES>
A:Cross-references: GB:S51940; NID:g262937; PIDN:AA24802.1; PID:g262938
C:Superfamily: herpesvirus complement control protein; complement factor H repeat homolog
F:35-94/Domain: complement factor H repeat homology <FH01>
F:99-157/Domain: complement factor H repeat homology <FH02>
F:162-223/Domain: complement factor H repeat homology <FH03>
F:228-283/Domain: complement factor H repeat homology <FH04>

Query Match 96.7%: Score 29; DB 2; Length 369;
Best Local Similarity 80.0%: Pred. No. 1.5e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
11 11
Db 11 FPMWR 15

RESULT 14
S26667
G protein-coupled receptor BLRI - human
C:Species: Homo sapiens (man)
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jul-2000
C:Accession: S26667
R:Dobner, T.; Wolff, I.; Emrich, T.; Lipp, M.
Eur. J. Immunol. 22: 2795-2799, 1992
A:Title: Differentiation-specific expression of a novel G protein-coupled receptor from A:Reference number: S26667; MUID:93049615
A:Accession: S26667
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-372 <DOB>
A:Cross-references: EMBL:X68149; NID:g29459; PIDN:CAA48252.1; PID:g29460
C:Genetics:
A:Gene: GDB:BLRI
A:Cross-references: GDB:136235; OMIM:601613

A:Map position: 15q26.1-15q26.1
C:Superfamily: vertebrate rhodopsin
C:Keywords: G protein-coupled receptor; transmembrane protein

Query Match 96.7%: Score 29; DB 2; Length 372;
Best Local Similarity 80.0%: Pred. No. 1.5e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
11 11
Db 352 FPMWR 356

RESULT 15
154479
membrane cofactor protein precursor, splice form pm5.1 - human
N:Alternate names: lymphocyte surface glycoprotein CD46
C:Species: Homo sapiens (man)
C>Date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 31-Mar-2000
C:Accession: 154479
S:Purcell, D.F.; Russell, S.M.; Deacon, N.J.; Brown, M.A.; Hooker, D.J.; McKenzie, I.
Immunogenetics 33: 335-344, 1991
A:Title: Alternatively spliced RNAs encode several isoforms of CD46 (MCP), a regulator of the immune response.
A:Reference number: 154479; MUID:91267562
A:Accession: 154479
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-377 <RES>
A:Cross-references: GB:M58050; NID:g180136; PIDN:AAA62833.1; PID:g180137
C:Genetics:
A:Gene: GDB:MCP
A:Cross-references: GDB:120169; OMIM:120920
A:Map position: 1q32-1q32
C:Function:
A:Description: for the factor I-mediated cleavage of the complement convertases
C:Superfamily: herpesvirus complement control protein; complement factor H repeat homolog
F:35-94/Domain: alternative splicing; glycoprotein; transmembrane protein
F:99-157/Domain: complement factor H repeat homology <FH01>
F:162-223/Domain: complement factor H repeat homology <FH02>
F:228-283/Domain: complement factor H repeat homology <FH03>
F:329-351/Domain: transmembrane #status predicted <TM>

Query Match 96.7%: Score 29; DB 2; Length 377;
Best Local Similarity 80.0%: Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
11 11
Db 11 FPMWR 15

Search completed: February 27, 2002, 11:52:58
Job time: 450 sec


```

DE DISULFIDE BOND FORMATION PROTEIN B (DISULFIDE OXIDOREDUCTASE)
DE (FRAGMENT).
CN DSB.
OS Vibrio alginolyticus.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=663;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=B138-2;
RA Nakamura T., Enomoto H., Unemoto T.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR DISULFIDE BOND FORMATION IN SOME
CC PERIPLASMIC PROTEINS. ACTS BY OXIDIZING THE DSB A PROTEIN (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE DSB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D83728; BAA12087.1; -
DR InterPro: IPR003752; DsbA.
DR Pfam: PF02600; DsbA.1.
KW Oxidoreductase; Redox-active center; Electron transport; Chaperone;
KW Transmembrane; Inner membrane.
FT DOMAIN 3 16 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 17 33 POTENTIAL.
FT DOMAIN 34 51 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 52 67 POTENTIAL.
FT DOMAIN 68 74 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 75 92 POTENTIAL.
FT DOMAIN 93 147 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 148 165 POTENTIAL.
FT DISULFID 43 46 REDOX-ACTIVE (BY SIMILARITY).
FT DISULFID 107 133 REDOX-ACTIVE (BY SIMILARITY).
FT NON_TER 165 165
SQ SEQUENCE 165 AA: 18745 MW: 43966 D5FAA3272D CRC64:

Query Match 96.7%; Score 29; DB 1; Length 165;
Best Local Similarity 80.0%; Pred. No. 37;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXWR 5
DB 113 FPSWR 117

RESULT 3
E2AA_ECOLI STANDARD: PRT: 259 AA.
AC P13810:
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HEAT-LABILE ENTEROTOXIN IIA, A CHAIN PRECURSOR (LT-IIA).
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN (1)
RP SEQUENCE FROM N.A.
RC MEDLINE=88032841; PubMed=2822667;
RA Pickett C.L., Weinstein D.L., Holmes R.K.;
RT Genetics of type IIA heat-labile enterotoxin of Escherichia coli:
RT operon fusions, nucleotide sequence, and hybridization studies.;
RL J. Bacteriol. 169:5180-5187(1987).

```

```

CC -1- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY
CC THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.
CC -1- SUBUNIT: HETEROHEXAMER OF ONE A CHAIN AND OF FIVE B CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M17894; AAA24093.1; -
DR PIR: A29831; A29831.
DR HSSP: P43528; ITIL.
DR InterPro: IPR001144; Enterotoxin_A.
DR Pfam: PF01375; Enterotoxin_A.1.
DR PRINTS: PR00771; ENTEROTOXIN_A.
KW Enterotoxin; Signal.
FT SIGNAL 1 18
FT CHAIN 19 259 HEAT-LABILE ENTEROTOXIN IIA, A CHAIN.
FT DISULFID 203 215 BY SIMILARITY.
FT ACT_SITE 128 128 BY SIMILARITY.
SQ SEQUENCE 259 AA: 29242 MW: 996531 A32CABBA CRC64:

Query Match 96.7%; Score 29; DB 1; Length 259;
Best Local Similarity 80.0%; Pred. No. 56;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXWR 5
DB 187 FPSWR 191

RESULT 4
GUNT_ERWCH STANDARD: PRT: 332 AA.
AC P27032;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE MINOR ENDOGLUCANASE Y PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE
DE Y) (CELLULOSE Y) (EGY).
CN CelY.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=556;
RN (1)
RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-33.
RC STRAIN=3937;
RC MEDLINE=92039050; PubMed=1937031;
RA Giuseppe A., Aymeric J.-L., Cami B., Barras F., Creuzet N.;
RT "Sequence analysis of the cellulase-encoding cell gene of Erwinia
RT chrysanthemi: a possible case of interspecies gene transfer.";
RL Gene 106:109-114(1991).
CC -1- FUNCTION: REPRESENTS ONLY 3% OF THE GLOBAL CELLULOSE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY D (FAMILY 8 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M74044; AAA24818.1; -
DR PIR: JT0585; JT0585.

```

DR InterPro: IPR002037; Glyco_hydro_8.
 DR Pfam: PF01270; Glyco_hydro_8; 1.
 DR PRINTS: PR00735; GLYCOLASE.
 DR PROSITE: PS00812; GLYCOSYL_HYDROL_F8; 1.
 KM Cellulose degradation: Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 23
 FT CHAIN 24 332
 FT ACT_SITE 53 53
 FT ACT_SITE 110 110
 SEQUENCE 332 AA: 37592 MW: 6EA760A2B8227079 CRC64;

Query Match: 96.7%; Score 29; DB 1; Length 332;
 Best Local Similarity 80.0%; Pred. No. 71;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FPKMR 5
 Db 178 FPMWR 182

RESULT 5
 CUN_CELUD STANDARD: PRT: 359 AA.
 AC P18336;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE) (CELLULOSE).
 OS Cellulomonas uda.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Micrococcales; Cellulomonadaceae; Cellulomonas.
 RX NCBI_TaxID=1714;
 RN [1]
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 24-37.
 RC STRAIN=CB4;
 RA Nakamura K., Misawa N., Kitamura K.;
 RT "Sequence of a cellulase gene of Cellulomonas uda CB4."
 RL J. Biotechnol. 4:247-254(1986).
 CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
 CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
 CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
 CC (2) EXOCELLULOBIOSACCHARIDASES THAT CUT THE DISACCHARIDE CELLULOSE
 CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
 CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER
 CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY D (FAMILY 8 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: M36503; AAA23090.1;
 DR InterPro: IPR002037; Glyco_hydro_8.
 DR Pfam: PF01270; Glyco_hydro_8; 1.
 DR PROSITE: PS00812; GLYCOSYL_HYDROL_F8; 1.
 KM Cellulose degradation: Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 23
 FT CHAIN 24 359
 FT ACT_SITE 53 53
 FT ACT_SITE 110 110
 SEQUENCE 359 AA: 40690 MW: 0445D571B6831A8 CRC64;

Query Match 96.7%; Score 29; DB 1; Length 359;

Best Local Similarity 80.0%; Pred. No. 76;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 FPKMR 5
 Db 178 FPMWR 182

RESULT 6
 CCR5_HUMAN STANDARD: PRT: 372 AA.
 AC P32302; 014811;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE C-X-C CHEMOKINE RECEPTOR TYPE 5 (CXCR-5) (CXCR-5) (BURKITT'S LYMPHOMA
 DE RECEPTOR 1) (MONOCYTE-DERIVED RECEPTOR 15) (MDR15).
 GN BLR1 OR CXCR5.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 RX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG FORM).
 RC TISSUE=Lymphocytes;
 RX MEDLINE=93049615; Pubmed=1425907;
 RA Doberer T., Wolf I., Emrich T., Lipp M.;
 RT "Differentiation-specific expression of a novel G protein-coupled
 RT receptor from Burkitt's lymphoma."
 RL Eur. J. Immunol. 22:2795-2799(1992).
 RN [2]
 RP SEQUENCE FROM N.A. (SHORT FORM).
 RC TISSUE=Blood;
 RX MEDLINE=95366951; Pubmed=7639692;
 RA Bazzella L., Loetscher M., Tobler A., Baggiolini M., Moser B.;
 RT "Sequence variation of a novel heptahelical leucocyte receptor
 RT through alternative transcript formation."
 RL Biochem. J. 309:773-779(1995).
 RN [3]
 RP LIGAND BINDING.
 RX MEDLINE=98130629; Pubmed=9463416;
 RA Legler D.F., Loetscher M., Stuber R., Clark-Lewis I.,
 RA Baggiolini M., Moser B.;
 RT "A cell-attracting chemokine 1, a human CXC chemokine expressed in
 RT lymphoid tissues, selectively attracts B lymphocytes via
 RT BLR1/CXCR5."
 RL J. Exp. Med. 187:655-660(1998).
 CC -1- FUNCTION: CYTOKINE RECEPTOR THAT BINDS TO BLC. BLR1 EXERTS
 CC POSSIBLY A REGULATORY FUNCTION IN BURKITT LYMPHOMA (BL)
 CC LYMPHOMAGENESIS AND/OR B-CELL DIFFERENTIATION. IT IS A POTENTIAL
 CC CANDIDATE FOR CELL-CELL INTERACTION, AND ACTIVATION OF MATURE B-
 CC LYMPHOCYTES IN LYMPHATIC TISSUES.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM: ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSION IN MATURE B-CELLS AND BURKITT
 CC LYMPHOMA CELLS.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: X68149; CAA48252.1;
 DR EMBL: X68829; CAA48723.1;
 DR PIR: S26667; S26667.
 DR CCRDb: CCR_0453;
 DR GCRDb: GCR_2072;
 DR GCRDb: GCR_2612;

DR MIM: 601613: -
 DR InterPro: IPR000276; GPCR_Rhodopsn.
 DR Pfam: PF00001; 7tm_1: 1.
 DR PRINTS: PR00237; GPCR_Rhodopsn.
 DR PRINTS: PR00564; BURKITTSLYMR.
 DR PROSITE: PS00237; G_PROTEIN_RECEP_F1_1: 1.
 DR PROSITE: PS00262; G_PROTEIN_RECEP_F1_3: 1.
 DR G-protein coupled receptor; Transmembrane; Glycoprotein; B-cell;
 KW Alternative splicing.
 FT TRANSMEM 1 55
 FT DOMAIN 56 76
 FT TRANSMEM 77 98
 FT TRANSMEM 89 109
 FT TRANSMEM 110 124
 FT TRANSMEM 125 145
 FT TRANSMEM 146 167
 FT TRANSMEM 168 188
 FT TRANSMEM 189 219
 FT TRANSMEM 220 240
 FT TRANSMEM 241 259
 FT TRANSMEM 260 280
 FT TRANSMEM 281 304
 FT TRANSMEM 305 325
 FT DOMAIN 326 372
 FT CARBOHYD 328 328
 FT CARBOHYD 196 196
 FT DISULFID 122 202
 FT VARSPLIC 1 45
 FT CONFLICT 344 344
 FT SEQUENCE 372 AA: 41955 MM: 6DF84C839492ACCF CRC64:
 Query Match 96.7%; Score 29; DB 1; Length 372;
 Best Local Similarity 80.0%; Pred. No. 79;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
 11 11
 Db 352 FPSWR 356

RESULT 7
 MCP_HUMAN STANDARD: PRT: 377 AA.
 ID MCP_HUMAN P15529;
 AC 01-APR-1990 (Rel. 14, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE MEMBRANE COFACTOR PROTEIN PRECURSOR (CD46 ANTIGEN) (TROPHOBLAST
 DE LEUCOCYTE COMMON ANTIGEN) (TLX).
 GN MCP.
 OS Homo sapiens (human).
 OC Rukayotla; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 OX 11
 RP SEQUENCE FROM N.A. AND SEQUENCE OF 35-58.
 RX MEDLINE=88286080; PubMed=3260937;
 RA Lublin D.M., Liszewski M.K., Post T.W., Arce M.A., le Beau M.M.,
 RA Rebenisch M.B., Lemons R.S., Seta T., Atkinson J.P.;
 RT "Molecular cloning and chromosomal localization of human membrane
 RT cofactor protein (MCP). Evidence for inclusion in the multigene
 RT family of complement-regulatory proteins.";
 RL J. Exp. Med. 168:181-194(1988).
 RN [12]
 RP SEQUENCE FROM N.A.
 RP TISSUE=Testis;
 RX MEDLINE=93119658; PubMed=8418811;
 RA Cervoni F., Fentchel P., Akhoundi C., Hsi B.L., Rossi B.;
 RT "Characterization of a cDNA clone coding for human testis membrane
 RT cofactor protein (MCP, CD46).";
 RL Mol. Reprod. Dev. 34:107-113(1993).
 RN [3]

RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
 KX MEDLINE=91267562; PubMed=2050389;
 RA Purcell D.F., Russell S.M., Deacon N.J., Brown M.A., Hooker D.J.,
 RA McKenzie I.F.;
 RT "Alternatively spliced RNAs encode several isoforms of CD46 (MCP), a
 RT regulator of complement activation.";
 RL Immunogenetics 33:335-344(1991).
 RN [4]
 RP SEQUENCE OF 1-34 FROM N.A.
 RX MEDLINE=94014356; PubMed=7691939;
 RA Cui W., Hourcade D., Post T., Greenlund A.C., Atkinson J.P.,
 RA Kumar V.;
 RT "Characterization of the promoter region of the membrane cofactor
 RT protein (CD46) gene of the human complement system and comparison to
 RT a membrane cofactor protein-like genetic element.";
 RL J. Immunol. 151:4137-4146(1993).
 RN [5]
 RP ALTERNATIVE SPLICING.
 KX MEDLINE=92289809; PubMed=1601037;
 RA Russell S.M., Sparrow R.U., McKenzie I.F.C., Purcell D.F.J.;
 RT "Tissue-specific and allelic expression of the complement regulator
 RT CD46 is controlled by alternative splicing.";
 RL Eur. J. Immunol. 22:1513-1518(1992).
 CC -1- FUNCTION: MCP IS AN IMPORTANT MEMBRANE PROTEIN FOR PROTECTING HOST
 CC CELLS FROM DAMAGE BY COMPLEMENT. MCP POSSESSES COFACTOR ACTIVITY
 CC FOR INACTIVATION OF C3B AND C4B BY SERUM FACTOR I. THIS WOULD
 CC ALLOW MCP TO REGULATE THE COMPLEMENT SYSTEM AT THE LEVEL OF THE C3
 CC CONVERTASES. MAY BE INVOLVED IN THE FUSION OF THE SPERMATOZOA WITH
 CC THE OOCYTE DURING FERTILIZATION. MAY ACT BY PREVENTING COMPLEMENT
 CC ACTIVATION AT THE FETUS-MATERNAL INTERFACE ON THE
 CC SYNCYTOTROPHOBLAST LAYER OF PLACENTA.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 14 FORMS ARE PRODUCED BY
 CC ALTERNATIVE SPLICING. THE FORM SHOWN IS THE TRANSCRIPT A. THE
 CC GAMMA ISOFORM (TRANSCRIPTS A AND B) IS PREFERENTIALLY EXPRESSED IN
 CC EMBRYO CELLS AND LEUKEMIC CELLS. THE 66 KDA ALPHA ISOFORM AND THE
 CC 56 KDA BETA ISOFORM FOUND IN ALL TISSUES EXCEPT SPERM CORRESPOND
 CC TO THE TRANSCRIPTS C AND D, AND THE TRANSCRIPTS E AND F
 CC RESPECTIVELY. THE EXON 9 IS SPECIFICALLY DELETED IN SOME
 CC PLACENTAE. THE EPSILON ISOFORM CORRESPONDS TO THE TRANSCRIPTS I
 CC AND J. SPERMATOZOON DELTA ISOFORM (TRANSCRIPT N) DELETE BOTH EXONS
 CC 12 AND 13. ALL TISSUES DIFFERENTIALLY SPLICED EXON 13.
 CC -1- TISSUE SPECIFICITY: EXPRESSED BY ALMOST ALL CELLS EXCEPT
 CC ERYTHROCYTES AND SOME BONE MARROW CELLS.
 CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED IN THE SER/THR-RICH DOMAIN
 CC (PROBABLE).
 CC -1- SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
 CC -1- DATABASE: NAME-PROW: NOTE-CD guide CD46 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd46.htm".
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Y00651; CAA68675.1; -;
 CC EMBL: S51940; AAB24802.1; -;
 CC EMBL: M58050; AAA62833.1; -;
 CC EMBL: A18585; CAA01400.1; -;
 CC EMBL: S65879; AAD19668.1; -;
 CC PIR: S01896; S01896.
 CC HSSP: P10998; 1VVC.
 DR MIM: 120920: -;
 DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF00084; sushi: 4.
 DR SMART: SM00032; CCP: 4.
 KW Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
 KM Sushi; Alternative splicing.
 FT SIGNAL 1 34
 FT CHAIN 35 377 MEMBRANE COFACTOR PROTEIN.

```

FT DOMAIN 35 328 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 329 351 POTENTIAL.
FT DOMAIN 352 377 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 35 95 SUSHI 1.
FT DOMAIN 98 158 SUSHI 2.
FT DOMAIN 161 224 SUSHI 3.
FT DOMAIN 227 284 SUSHI 4.
FT DOMAIN 287 311 SER/THR-RICH.
FT DISULFID 35 80 BY SIMILARITY.
FT DISULFID 64 94 BY SIMILARITY.
FT DISULFID 99 141 BY SIMILARITY.
FT DISULFID 127 157 BY SIMILARITY.
FT DISULFID 162 210 BY SIMILARITY.
FT DISULFID 191 223 BY SIMILARITY.
FT DISULFID 228 270 BY SIMILARITY.
FT CARBOHYD 236 283 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 273 273 MISSING (IN ISOFORM N.).
FT VARSPPLIC 286 300 MISSING (IN A SECOND FORM).
FT VARSPPLIC 340 352 MISSING (IN ISOFORM M).
FT VARSPPLIC 353 361 YLDKRRKKG -> DIFGGRKKGQVVELMPLTRLNPLQ
FT VARSPPLIC 362 377 QSRRAE (IN ISOFORM M).
FT VARSPPLIC 362 377 TYLDETHREVEKFTSL -> KADGAEVATYQTKSTPAEO
FT VARSPPLIC 362 377 RG (IN ISOFORM B/D/E/H/J/L).
FT VARSPPLIC 340 361 MISSING (IN ISOFORM M AND ISOFORM N).
FT VARSPPLIC 339 339 MISSING (IN ISOFORM N).
FT VARSPPLIC 339 339 I -> IGKQVELMPLTRLNPLQSQSRRAE (IN
SQ SEQUENCE 377 AA: 42247 MW: 2CA6F61752570B57 CRC64:
  Query Match 96.7%; Score 29; DB 1; Length 377;
  Best Local Similarity 80.0%; Pred. No. 80;
  Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
  QY 1 FPXWR 5
  DB 11 FPSMR 15
  ID 2229 HUMAN STANDARD; PRT: 420 AA.
  AC Q9UWJ7;
  DT 20-AUG-2001 (Rel. 40, Created)
  DT 20-AUG-2001 (Rel. 40, Last sequence update)
  DT 20-AUG-2001 (Rel. 40, Last annotation update)
  DE ZINC FINGER PROTEIN 229 (FRAGMENT).
  GN ZNF229.
  OS Homo sapiens (Human).
  OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
  OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
  OX NCBI_TaxID=9606;
  RN [1]
  RP SEQUENCE FROM N.A.
  RA Shannon M., Branscomb E., Hauser L., Gordon L., Ashworth L.,
  RA Stubbs L.;
  RT "Differential expansion of homologous zinc-finger gene families in
  RT human chromosome 19q13.2 and mouse chromosome 7.";
  RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
  RN [2]
  RP SEQUENCE FROM N.A.
  RA Kodoyanli V., Ge Y., Krummel G.K., Kvistad E., Grable L.,
  RA Severin J., Gordon L., Shannon M., Brower A., Olsen A.S., Smith L.M.;
  RT "Sequence analysis of a 1Mb region in 19q13.2 containing a zinc finger
  RT gene cluster.";
  RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
  CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
  CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
  CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
  CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF192979; AAF07964.1; -
DR EMBL: AC084239; AAG23970.1; -
DR InterPro: IPR001909; KRAB.
DR InterPro: IPR008822; ZnF-C2H2.
DR Pfam: PF01352; KRAB.1.
DR Pfam: PF00096; zf-C2H2.2.
DR PRINTS: PR00048; ZNCFINGER.
DR SMART: SM00349; KRAB.1.
DR SMART: SM00355; ZnF-C2H2.2.
DR PROSITE: PS00805; KRAB.1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 2.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 3.
DR Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 34 108 KRAB.
FT DOMAIN 349 >420 ZINC_FINGERS.
FT ZN_FING 349 371 C2H2-TYPE.
FT ZN_FING 377 399 C2H2-TYPE.
FT ZN_FING 405 >420 C2H2-TYPE.
FT NON_TER 420 420
SQ SEQUENCE 420 AA: 48022 MW: FA4138BA4A4A2A14 CRC64:
  Query Match 96.7%; Score 29; DB 1; Length 420;
  Best Local Similarity 80.0%; Pred. No. 89;
  Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
  QY 1 FPXWR 5
  DB 173 FPAWR 177
  ID HS16 CAEEL STANDARD; PRT: 143 AA.
  AC P06581; P02514;
  DT 01-JAN-1988 (Rel. 06, Created)
  DT 01-JAN-1988 (Rel. 06, Last sequence update)
  DT 01-FEB-1996 (Rel. 33, Last annotation update)
  DE HEAT SHOCK PROTEIN HSP16-41.
  GN HSP16-41.
  OS Caenorhabditis elegans.
  OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
  OC Rhabditidae; Pelodierinae; Caenorhabditis.
  OX NCBI_TaxID=6239;
  RN [1]
  RP SEQUENCE FROM N.A.
  RA MEDLINE=86304344; PubMed=1017958;
  RA Jones D., Russnak R.H., Kay R.J., Candido E.P.M.;
  RT "Structure, expression, and evolution of a heat shock gene locus in
  RT Caenorhabditis elegans that is flanked by repetitive elements.";
  RL J. Biol. Chem. 261:12006-12015(1986).
  RN [2]
  RP SEQUENCE OF 47-143 FROM N.A.
  RA MEDLINE=83220736; PubMed=6190129;
  RA Russnak R.H., Jones D., Candido E.P.M.;
  RT "Cloning and analysis of cDNA sequences coding for two 16 kilodalton
  RT heat shock proteins (hsps) in Caenorhabditis elegans: homology with
  RT the small hsps of Drosophila.";
  RL Nucleic Acids Res. 11:3187-3205(1983).
  CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
  CC FAMILY.
  CC -----
  CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: M14334; AAA28070.1; ALT-SEQ.
 DR EMBL: X01577; CAA25732.1; -.
 DR PIR: A38884; HHKWL1.
 DR PIR: A25199; A25199.
 DR InterPro: IPR002068; Crystallin_HSP20.
 DR Pfam: PF00041; HSP20; 1.
 DR PROSITE: PS01031; HSP20; 1.
 KW Heat shock; Multigene family.
 KM SEQUENCE 143 AA; 16252 MW; C1D0F59D26E36C24 CRC64;
 SQ

Query Match 93.3%; Score 28; DB 1; Length 143;
 Best Local Similarity 80.0%; Pred. No. 50;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXWR 5
 11 11
 Db 27 FPYWR 31

RESULT 10
 ID HS17_CAEEL STANDARD: PRT: 143 AA.
 AC P02513;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE HEAT SHOCK PROTEIN HSP16-48.
 GN HSP16-48.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN 11
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85295957; PubMed=4033652;
 RA Russnak R.H., Candido E.P.M.;
 RT "Locus encoding a family of small heat shock genes in Caenorhabditis
 RT elegans: two genes duplicated to form a 3.8-kilobase inverted
 RT repeat";
 RL Mol. Cell. Biol. 5:1268-1276(1985).
 RN 121
 RP SEQUENCE OF 9-143 FROM N.A.
 RX MEDLINE=83220736; PubMed=6190129;
 RA Russnak R.H., Jones D., Candido E.P.M.;
 RT "Cloning and analysis of cDNA sequences coding for two 16 kilodalton
 RT heat shock proteins (hsps) in Caenorhabditis elegans: homology with
 RT the small hsps of Drosophila";
 RL Nucleic Acids Res. 11:3187-3205(1983).
 CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: K03273; AAA28069.1; -.
 DR EMBL: X01576; CAA25731.1; -.
 DR EMBL: K01863; AAA28064.1; -.
 DR PIR: A02916; HHKWL8.
 DR PIR: A24289; A24289.
 DR InterPro: IPR002068; Crystallin_HSP20.

DR Pfam: PF00011; HSP20; 1.
 DR PROSITE: PS01031; HSP20; 1.
 KW Heat shock; Multigene family.
 SQ SEQUENCE 143 AA; 16299 MW; 0D5596DFEE5B3318 CRC64;

Query Match 93.3%; Score 28; DB 1; Length 143;
 Best Local Similarity 80.0%; Pred. No. 50;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXWR 5
 11 11
 Db 27 FPYWR 31

RESULT 11
 ID HNL_MANES STANDARD: PRT: 257 AA.
 AC P52705;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE (S)-ACETONE-CYANOHYDRIN LYASE (EC 4.1.2.39) ((S)-HYDROXYNITRILE LYASE)
 DE ((S)-HYDROXYNITRILASE) (OXNITRILASE).
 GN HNL.
 OS Manihot esculenta (Cassava) (Manioc).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids I; Malpighiales; Euphorbiaceae; Manihot.
 OX NCBI_TaxID=3983;
 RN 111
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-36 AND 168-191.
 RC TISSUE=Cotyledon;
 RX MEDLINE=94263231; PubMed=8203915;
 RA Hughes J., Carvalho F.J.P.D.C., Hughes M.A.;
 RT "Purification, characterization, and cloning of alpha-hydroxynitrile
 RT lyase from cassava (Manihot esculenta Gratz)";
 RL Arch. Biochem. Biophys. 311:496-502(1994).
 CC -1- FUNCTION: INVOLVED IN CYANOCENESIS. THE RELEASE OF HCN FROM
 CC INJURED TISSUES. DECOMPOSES A VARIETIES OF (R) OR (S) CYNOHYDRINS
 CC INTO HCN AND THE CORRESPONDING ALDEHYDES AND KETONES. THE NATURAL
 CC SUBSTRATE OF THIS ENZYME IS (S)-ACETONE CYANOHYDRIN.
 CC -1- CATALYTIC ACTIVITY: (S)-2-HYDROXYISOBUTYRONITRILE = CYANIDE +
 CC ACETONE.
 CC -1- SUBUNIT: HOMOTRIMER.
 CC -1- SIMILARITY: DISTANT RELATIONSHIP WITH TYPE-B
 CC CARBOXYL-ESTERASE/LIPASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: 229091; CAA82334.1; -.
 DR HSSP: P52704; LYAS.
 DR InterPro: IPR000073; Abhydrolase.
 DR InterPro: IPR000379; Est_lip_thioest_actsite.
 DR Pfam: PF00561; abhydrolase; 1.
 KW Lyase.
 FT INIT_MET 0
 FT ACT_SITE 79 79 BY SIMILARITY.
 FT ACT_SITE 207 207 BY SIMILARITY.
 FT ACT_SITE 235 235 BY SIMILARITY.
 ET SEQUENCE 257 AA; 29240 MW; 98B3E160ACB33BC5 CRC64;

Query Match 93.3%; Score 28; DB 1; Length 257;
 Best Local Similarity 80.0%; Pred. No. 87;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;


```

OY      1 1 PPXMR 5
DB      124 FPDWR 128

RESULT 12
PABC_VIBHA STANDARD: PRT: 271 AA.
ID PABC_VIBHA 056693:
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE 4-AMINO-4-DEOXYCHORISMATE LYASE (EC 4.1.1.1) (ADC LYASE) (ADCL).
GN PABC.
OS Vibrio harveyi.
OC Bacteria: Proteobacteria: gamma subdivision: Vibrionaceae: Vibrio.
OX NCBI_TaxID=669;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B392;
RX MEDLINE=96134997; PubMed=8550484;
RA Shen Z., Byers D.M.;
RT "Isolation of Vibrio harveyi acyl carrier protein and the fabG, accP,
RL J. Bacteriol. 178:571-573(1996).
CC -1- FUNCTION: CONVERTS 4-AMINO-4-DEOXYCHORISMATE INTO 4-AMINO-BENZOATE
(PABA) AND PYRUVATE.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -1- PATHWAY: SECOND STEP IN FOLATE BIOSYNTHESIS PATHWAY.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT
AMINOTRANSFERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U39441; AAC43592.1; -
DR InterPro: IPR001544; AminoTran_4.
DR Pfam: PF01063; aminoTran_4; 1.
DR ProDom: PD001961; AminoTran_4; 1.
DR PROSITE: PS00770; AA_TRANSFERRASE CLASS 4; 1.
KM Lyase: Pyridoxal phosphate: folate biosynthesis. (BY SIMILARITY).
FT BINDING 140 140 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 271 AA; 29958 MW; 3AE58B9F398D1CD CRC64;

Query Match 93.3%; Score 28; DB 1; Length 271;
Best Local Similarity 80.0%; Pred. No. 91;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1 1 PPXMR 5
DB      57 FPDWR 61

RESULT 13
PPNK_BACST STANDARD: PRT: 271 AA.
ID PPNK_BACST 558055:
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE PROBABLE INORGANIC POLYPHOSPHATE/ATP-NAD KINASE (EC 2.7.1.23)
DE (POLY(P)/ATP NAD KINASE).
GN PPNK OR NADK.
OS Bacillus stearothermophilus.
OC Bacteria: Firmicutes: Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Geobacillus.

```

```

OX NCBI_TaxID=1122;
RN [1]
RP SEQUENCE FROM N.A.
RA Oshima T., Sakuraba H.;
RT "Inorganic polyphosphate/ATP-NAD kinase of Bacillus
RN stearothermophilus";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CATALYZES THE PHOSPHORYLATION OF NAD TO NADP. UTILIZES
CC ATP AND OTHER NUCLEOSIDE TRIPHOSPHATES AS WELL AS INORGANIC
CC POLYPHOSPHATE AS A SOURCE OF PHOSPHORUS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + NAD(+) -> ADP + NADP(+)
CC -1- COFACTOR: REQUIRES DIVALENT METAL IONS FOR ACTIVITY (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NAD KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB055961; BAB32727.1; -
DR Transferase: Kinase; NAD: NADP.
KM Transferase: Kinase; NAD: NADP.
SQ SEQUENCE 271 AA; 30517 MW; 9E0FB427BC76662A CRC64;

Query Match 93.3%; Score 28; DB 1; Length 271;
Best Local Similarity 80.0%; Pred. No. 91;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1 1 PPXMR 5
DB      258 FPDWR 262

RESULT 14
PAL_KLEPN STANDARD: PRT: 286 AA.
ID PAL_KLEPN
AC P37446;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PHOSPHOLIPASE A1 PRECURSOR (EC 3.1.1.32) (DETEGENT-RESISTANT
DE PHOSPHOLIPASE A) (DR-PHOSPHOLIPASE A) (PHOSPHATIDYLCHOLINE 1-
DE ACYLHYDROLASE) (OUTER MEMBRANE PHOSPHOLIPASE A) (OM PLA).
GN PLDA.
OS Klebsiella pneumoniae.
OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94131966; PubMed=8300539;
RA Brock R.G.P.M., Brinkman E., van Boxtel R., Bekkers A.C.A.P.,
RA Verheij H.M., Tommassen J.;
RT "Molecular characterization of enterobacterial plda genes encoding
RT outer membrane phospholipase A.";
RL J. Bacteriol. 176:861-870(1994).
CC -1- FUNCTION: HYDROLYSIS OF PHOSPHATIDYLCHOLINE WITH PHOSPHOLIPASE
CC A2 (EC 3.1.1.4) AND PHOSPHOLIPASE A1 (EC 3.1.1.32) ACTIVITIES.
CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O -> 1-ACYLGLYCERO-
CC PHOSPHOCHOLINE + A FATTY ACID ANION.
CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O -> 2-ACYLGLYCERO-
CC PHOSPHOCHOLINE + A FATTY ACID ANION.
CC -1- COFACTOR: REQUIRES CALCIUM IONS FOR ACTIVITY.
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE; ONE OF THE VERY FEW ENZYMES
CC LOCATED THERE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X76901: CAA54223.1: -
 DR PIR: B36971: B36971.
 DR PIR: S40129: S40129.
 DR InterPro: IPR003187: PLAI.
 DR Pfam: PF02253: PLAI: 1.
 KM Hydrolysis: Lipid degradation: Outer membrane: Signal: Calcium.
 FT SIGMUL: 1 20 BY SIMILARITY
 FT CHAIN: 21 286 PHOSPHOLIPASE A1.
 FT ACT_SITE: 161 161 BY SIMILARITY
 SQ SEQUENCE 286 AA: 3E39F863085108A3 CRC64:

Query Match 93.3% Score 28; DB 1: Length 286;
 Best Local Similarity 80.0% Pred. No. 96;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 PPMWR 5
 Db 92 PPLMR 96

RESULT 15
 PAL_ECOLI STANDARD: PRT: 289 AA.
 AC P00631:
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PHOSPHOLIPASE A1 PRECURSOR (EC 3.1.1.32) (DETERGENT-RESISTANT
 DE PHOSPHOLIPASE A) (DR-PHOSPHOLIPASE A) (PHOSPHATIDYLCHOLINE 1-
 DE ACYLHYDROLASE) (OUTER MEMBRANE PHOSPHOLIPASE A) (OM PLA).
 CN PIDA OR B3821 OR Z5342 OR ECS4751.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562, 83334:
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85157492; PubMed=6397464;
 RA Homma H., Kobayashi T., Chiba N., Karasawa K., Mizushima H., Kudo I.,
 RA Inoue K., Ikeda H., Sekiguchi M., Nojima S.:
 RT "The DNA sequence encoding pida gene, the structural gene for
 RT detergent-resistant phospholipase A of E. coli.";
 RL J. Biochem. 96:1655-1664(1984).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX STRAIN-K12 / MG1655;
 RX MEDLINE=92358234; PubMed=1379743:
 RA Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.:
 RT "Analysis of the Escherichia coli genome: DNA sequence of the region
 RT from 84.5 to 86.5 minutes.";
 RL Science 257:771-778(1992).
 RN [3]
 RP REVISION TO 14-15
 RX STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.:
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RP SOURCE/NCBI FROM N.A.
 RX STRAIN-O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postell G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Goebbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.:
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533(2001).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX STRAIN-O157:H7 / RMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Martino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shingawa H.:
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 RN [6]
 RP SEQUENCE OF 30-289 FROM N.A.
 RX STRAIN-K12;
 RX MEDLINE=85003590; PubMed=6383820;
 RA de Gans P., Verheij H.M., Riegman N.H., Hoekstra W.P.M., de Haas G.H.:
 RT "The pro- and mature forms of the E. coli K-12 outer membrane
 RT phospholipase A are identical.";
 RL EMBO J. 3:1799-1802(1984).
 RN [7]
 RP SEQUENCE OF 174-289 FROM N.A.
 RX STRAIN-K12;
 RX MEDLINE=87115164; PubMed=3027506;
 RA Irlino N., Nakayama K., Nakayama H.:
 RT "The recG gene of Escherichia coli K12: primary structure and
 RT evidence for SOS regulation.";
 RL Mol. Gen. Genet. 205:298-304(1986).
 RN [8]
 RP MUTAGENESIS OF SER-172.
 RX MEDLINE=94131966; PubMed=8300539;
 RA Brok R.G.P.M., Brinkman E., van Boxtel R., Bekkers A.C.A.P.,
 RA Verheij H.M., Tommassen J.:
 RT "Molecular characterization of enterobacterial pida genes encoding
 RT outer membrane phospholipase A.";
 RL J. Bacteriol. 176:861-870(1994).
 RN [9]
 RP ACTIVE SITE SER-164, AND PARTIAL SEQUENCE.
 RX MEDLINE=91249806; PubMed=2040286;
 RA Horrevorts A.J.G., Verheij H.M., de Haas G.H.:
 RT "Inactivation of Escherichia coli outer-membrane phospholipase A by
 RT the affinity label hexadecanesulfonyl fluoride. Evidence for an
 RT active-site serine.";
 RL Eur. J. Biochem. 198:247-253(1991).
 CC -1- FUNCTION: HYDROLYSIS OF PHOSPHATIDYLCHOLINE WITH PHOSPHOLIPASE
 CC A2 (EC 3.1.1.4) AND PHOSPHOLIPASE A1 (EC 3.1.1.32) ACTIVITIES.
 CC REQUIRING FOR EFFICIENT SECRETION OF BACTERIOICINS, SEEMS TO BE
 CC DOMINANT IN NORMAL GROWING CELLS.
 CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O -> 1-ACYLGLYCERO-
 CC PHOSPHOCHOLINE + A FATTY ACID ANION.
 CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O -> 2-ACYLGLYCERO-
 CC PHOSPHOCHOLINE + A FATTY ACID ANION.
 CC -1- COFACTOR: REQUIRES CALCIUM IONS FOR ACTIVITY.
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE; ONE OF THE VERY FEW ENZYMES
 CC LOCATED THERE.
 CC -1- INDUCTION: BY MEMBRANE DAMAGING, FOR EXAMPLE, BY PHAGE-INDUCED
 CC LYSIS OR TEMPERATURE SHOCK.
 CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X02143: CAA26081.1; -.
DR EMBL: M87049: AAA67617.1; -.
DR EMBL: AE000458: AAC76824.1; -.
DR EMBL: AE005613: AAG59017.1; -.
DR EMBL: AP002567: BAB38174.1; -.
DR EMBL: M30198: AAA24516.1; -.
DR PIR: A00771: PSECA.
DR PIR: A2133: PSECAI.
DR PIR: S30711: S30711.
DR EcoGene: EG10738: PldA.
DR InterPro: IPR003187: PLA1.
DR Pfam: PF02253: PLA1; 1.
KW Hydroxylase; Lipid degradation; Outer membrane; Signal; Calcium;
Complete proteome.
KW SIGNAL 1 20
FT CHAIN 21 289 PHOSPHOLIPASE A1.
FT ACT_SITE 164 164
FT MUTAGEN 172 172 S->F: INACTIVE PROTEIN.
FT CONFLICT 14 15 LP -> FA (IN REF. 2).
FT CONFLICT 30 33 DAPA -> MTRQ (IN REF. 6).
SQ SEQUENCE 289 AA: 33163 MW: A688AD32AA60F218 CRC64;

Query Match 93.38; Score 28; DB 1; Length 289;
Best Local Similarity 80.08; Pred. No. 97;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PPXNR 5
DB 95 PPLWR 99

Search completed: February 27, 2002, 11:52:17
Job time: 564 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:54:08 ; Search time 62.61 Seconds
(without alignments)
11.681 Million cell updates/sec

Title: US-09-446-109a-22
Sequence: 1 FPXWR 5

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP_mhc:*
8: SP organelle:*
9: SP-phage:*
10: SP-plant:*
11: SP-rodent:*
12: SP-virus:*
13: SP-vertebrate:*
14: SP-unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	96.7	101	5 Q9V671	Q9V671 drosophila
2	29	96.7	113	10 Q39293	Q39293 brassica na
3	29	96.7	157	5 Q9X7Y5	Q9X7Y5 caenorhabdi
4	29	96.7	158	4 Q9NNW2	Q9NNW2 homo sapien
5	29	96.7	158	4 Q9NNW2	Q9NNW2 homo sapien
6	29	96.7	229	8 Q37633	Q37633 riagoletis
7	29	96.7	245	2 Q9PDE1	Q9PDE1 xyliella tas
8	29	96.7	273	10 Q9SUV9	Q9SUV9 arabidopsis
9	29	96.7	292	2 Q9H7B3	Q9H7B3 pseudomonas
10	29	96.7	321	11 Q9D6L7	Q9D6L7 mus musculu
11	29	96.7	332	2 Q9APJ5	Q9APJ5 erwinia chr
12	29	96.7	333	2 Q9L3G9	Q9L3G9 erwinia rha
13	29	96.7	334	10 Q9FTW4	Q9FTW4 oryza sativ
14	29	96.7	349	4 Q15429	Q15429 homo sapien
15	29	96.7	417	10 Q9SA41	Q9SA41 arabidopsis
16	29	96.7	466	2 Q9FUK4	Q9FUK4 zymomonas m
17	29	96.7	479	2 Q86742	Q86742 streptomyces
18	29	96.7	497	2 Q91702	Q91702 pseudomonas
19	29	96.7	501	5 Q16923	Q16923 caenorhabdi

20	29	96.7	502	5 Q9GUIS	Q9GUIS caenorhabdi
21	29	96.7	555	1 Q29748	Q29748 archaeoglob
22	29	96.7	581	4 Q9BR01	Q9BR01 homo sapien
23	29	96.7	612	4 Q9HB92	Q9HB92 homo sapien
24	29	96.7	621	4 Q9H9Y1	Q9H9Y1 homo sapien
25	29	96.7	625	2 Q66676	Q66676 aquifex aeo
26	29	96.7	634	4 Q9NM19	Q9NM19 homo sapien
27	29	96.7	634	4 Q9NM19	Q9NM19 homo sapien
28	29	96.7	642	2 Q25396	Q25396 helicobacte
29	29	96.7	650	5 Q17866	Q17866 caenorhabdi
30	29	96.7	735	2 Q9H7B3	Q9H7B3 pseudomonas
31	29	96.7	778	5 Q9U9K6	Q9U9K6 caenorhabdi
32	29	96.7	811	4 Q9BWX2	Q9BWX2 homo sapien
33	29	96.7	1071	11 Q9D2K4	Q9D2K4 mus musculu
34	29	96.7	1108	13 Q9PMD0	Q9PMD0 tetraodon f
35	29	96.7	1117	5 Q9U9K7	Q9U9K7 caenorhabdi
36	29	96.7	1451	2 Q45336	Q45336 bordetella
37	29	96.7	1451	2 Q45044	Q45044 bordetella
38	29	96.7	1451	2 Q9S5D5	Q9S5D5 bordetella
39	29	96.7	168	2 P73858	P73858 synechocyst
40	28	93.3	84	2 P74471	P74471 synechocyst
41	28	93.3	84	2 P74471	P74471 synechocyst
42	28	93.3	84	2 P73028	P73028 synechocyst
43	28	93.3	93	2 P73028	P73028 synechocyst
44	28	93.3	103	10 Q9M045	Q9M045 arabidopsis
45	28	93.3	110	6 Q9GLD9	Q9GLD9 canis famill

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	101 AA.
Q9V671	Q9V671			
AC	Q9V671			
UF	01-MAY-2000 (TREMBLrel. 13, Created)			
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	CG13174 PROTEIN.			
GN	CG13174.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
CC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
CC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
SN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BERKELEY:			
RA	MEDLINE-20196006; PubMed-10731132;			
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,			
RA	George R.A., Lewis S.F., Richards S., Ashburner M., Henderson S.N.,			
RA	Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,			
RA	Branton R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,			
RA	Wan K.H., Doyle G., Baxter F.G., Helt G., Nelson C.R., Miklos G.L.G.,			
RA	Abail J.F., Agbayani A., An H.-J., Andrews-Piankovich C., Baldwin D.,			
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,			
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,			
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,			
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,			
RA	Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,			
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			
RA	Dodson K., Dou P.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,			
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferricra S., Fleischmann W.,			
RA	Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,			
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,			
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,			
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,			
RA	Jajali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,			
RA	Kimmel P.E., Kodira C.U., Kraft C., Kravitz S., Kulp D., Lai Z.,			
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,			
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,			
RA	Meikunov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,			

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Sinden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svrtskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Morley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003823; AAF58560.1; -
 DR FlyBase: FBgn0033694; CG13174.
 SO SEQUENCE 101 AA; 11820 MW; D5C4D85B514CB9B8 CRC64;

Query Match 96.7%: Score 29; DB 5; Length 101;
 Best Local Similarity 80.0%: Pred. No. 1.2e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FPMXR 5
 Db 75 FPMXR 79

RESULT 2
 ID 039293 PRELIMINARY; PRT; 113 AA.
 AC 039293;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE ELONGATION FACTOR EF-1A (FRAGMENT).
 OS Brassica napus (Rape).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustosids II; Brassicales; Brassicaceae; Brassica.
 OX NCBI_TaxID=3708;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SAMOUBAI; TISSUE=ETIOLATED SEEDLINGS. (5-DAYS-OLD);
 RA Saez-Vasquez J.;
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 CC -I- SIMILARITY: TO GTP-BINDING ELONGATION FACTOR FAMILY.
 DR EMBL: U21744; AAA86366.1; -
 DR Mendel: 15767; Brana:1139;15767.
 DR InterPro: IPR000795; GTP_EFTU.
 DR Pfam: PF00009; GTP_EFTU; 1.
 KW Elongation factor; GTP-binding; Protein biosynthesis.
 FT NON_TER 1
 SO SEQUENCE 113 AA; 12560 MW; 149783D707223948 CRC64;

Query Match 96.7%: Score 29; DB 10; Length 113;
 Best Local Similarity 80.0%: Pred. No. 1.4e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FPMXR 5
 Db 103 FPMXR 107

RESULT 3
 ID 09XTY5 PRELIMINARY; PRT; 157 AA.
 AC 09XTY5;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE T03D8.2 PROTEIN.

GN T03D8.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Mortimore B.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Riken L., Koopra A., Saunders D., Showkhen R.,
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wollman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*."
 RL Nature 368:32-38(1994).
 DR EMBL: Z92838; CAB07406.1; -
 DR InterPro: IPR000230; Ribosomal_S12.
 DR Pfam: PF00164; Ribosomal_S12; 1.
 DR PRINTS: PR01034; RIBOSOMALS12.
 DR ProDom: PD000576; Ribosomal_S12; 1.
 SO SEQUENCE 157 AA; 17370 MW; 6C6A837282EB26C CRC64;

Query Match 96.7%: Score 29; DB 5; Length 157;
 Best Local Similarity 80.0%: Pred. No. 1.9e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FPMXR 5
 Db 40 FPMXR 44

RESULT 4
 ID 09NNW3 PRELIMINARY; PRT; 158 AA.
 AC 09NNW3;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE MEMBRANE COFACTOR PROTEIN CD46 VARIANT (FRAGMENT).
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eulheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20227770; PubMed=10751138;
 RA Kusuhara K., Sasaki Y., Nakao F., Ihara K., Hattori H., Yamashita S.,
 RA Nihel K., Koide N., Alba H., Takeshita K., Hara T.;
 RT "Analysis of measles virus binding sites of the CD46 gene in patients
 RT with subacute sclerosing panencephalitis."
 SL J. Infect. Dis. 181:1447-1449(2000).
 DR EMBL: AF209713; AAF73845.1; -
 DR InterPro: IPR000436; Sush1_SCR_CCP.
 DR Pfam: PF00084; sush1; 2.
 DR SMART: SM00032; CCP; 2.
 DR VARIANT 59
 FT NON_TER 59
 SO SEQUENCE 158 AA; 18098 MW; 4E3F07EEC5C454F5 CRC64;

Query Match 96.7%: Score 29; DB 4; Length 158;

```

Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
    1 1 1
Db 11 FPSMR 15

RESULT 5
O9NMW2 PRELIMINARY: PRT: 158 AA.
ID O9NMW2
AC O9NMW2: 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE MEMBRANE COPFACTOR PROTEIN CD46 VARIANT (FRAGMENT).
CN MCP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20227770; PubMed=10751138;
RA Kusuhara K., Sasaki Y., Nakao F., Ihara K., Hattori H., Yamashita S.,
RA Nihel K., Koide N., Alta H., Takeshita K., Hara T.;
RT "Analysis of measles virus binding sites of the CD46 gene in patients
RT with subacute sclerosing panencephalitis.";
RL J. Infect. Dis. 181:1447-1449(2000).
DR EMBL: AF209714; AAF73846.1;
DR InterPro: IPR000436; Sushl_SCR_CCP.
DR Pfam: PF00084; sushl; 2.
DR SMART: SM00032; CCP; 2.
DR NON_TER 158
FT SEQUENCE 158 AA; 18126 MW; E2CF38EC3A16A9DA CRC64;
SO

Query Match 96.7%; Score 29; DB 4; Length 158;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
    1 1 1
Db 11 FPSMR 15

RESULT 6
O37633 PRELIMINARY: PRT: 229 AA.
ID O37633
AC O37633: 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE II (EC 1.9.3.1) (FRAGMENT).
CN COLL.
OS Rhagoletis boycell.
OC Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Tephritidae; Tephritidae; Rhagoletis.
OX NCBI_TaxID=43419;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97159559; PubMed=9007018;
RA Smith J.J., Bush G.L.;
RT "Phylogeny of the genus Rhagoletis (Diptera: Tephritidae) inferred
RT from DNA sequences of mitochondrial cytochrome oxidase II.";
RL Mol. Phylogenet. Evol. 7:33-43(1997).
DE -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
DE CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
DE 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
DE TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
DE A CENTER TO THE BINUCLEAR CENTER OF THE CATALYTIC SUBUNIT 1 (BY

```

```

CC SIMILARITY)
CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O + 4
CC FERRICYTOCHROME C.
CC -1- COPFACTOR: COPPER A AND HEME GROUP.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL.
CC INNER MEMBRANE (BY SIMILARITY).
CC -1- SIMILARITY: TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
DR EMBL: U53254; AAB50355.1;
DR HSSP: P08306; IARI.
DR InterPro: IPR001505; COX2.
DR InterPro: IPR002429; Cyt_cox_2.
DR Pfam: PF00116; COX2; 1.
DR PRINTS: PR01166; CYCOXIDASEII.
DR PRODOM: PD000131; COX2; 1.
DR PROSITE: PS00078; COX2; 1.
KW Copper; Inner membrane; Mitochondrion; Oxidoreductase;
KW Respiratory chain; Transmembrane.
FT METAL 159
FT METAL 163
FT METAL 167
FT METAL 207
FT METAL 207
FT NON_TER 229
FT SEQUENCE 229 AA; 26389 MW; 912D1B2736330698 CRC64;
SO

Query Match 96.7%; Score 29; DB 8; Length 229;
Best Local Similarity 80.0%; Pred. No. 2.7e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
    1 1 1
Db 78 FPSMR 82

RESULT 7
O9PDEL PRELIMINARY: PRT: 245 AA.
ID O9PDEL
AC O9PDEL: 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE HYPOTHEICAL PROTEIN XF1438.
CN XF1438.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Rehnach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Bala J.M., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.S., Brlones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carreir H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohne M., Furlan L.R.,
RA Carrier M., Goldman C.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohnselt J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lamais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Martino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorillo C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhai A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmeri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.F.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,

```

RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zaitz M., Meidanis J., Setubal J.C.;
 RT "the genome sequence of the plant pathogen *Xylella fastidiosa*.";
 RL Nature 406:151-159(2000).
 DR EMBL: AE003974; AAF84247.1; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 245 AA; 27193 MW; 16CBAD9B9EC490455 CRC64;

Query Match: 96.7%; Score 29; DB 2; Length 245;
 Best Local Similarity 80.0%; Pred. No. 2.8e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 FPMWR 5
 Db 197 FPMWR 201

RESULT 8
 ID 09SUV9 PRELIMINARY: PRT: 273 AA.
 AC 09SUV9;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE HYPOHETICAL 31.2 KDA PROTEIN.
 CN FPK2.180 OR AT4G22600.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Medler H., Wambutt R., Bancroft I., Mewes H.W.,
 RA Mayer K.F.X., Lemcke K., Schueller C.;
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Medler H., Wambutt R., Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL033545; CAA22165.1; -;
 DR EMBL: AL161557; CAB79215.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 273 AA; 31248 MW; 7E644F68CEB8B0 CRC64;

Query Match: 96.7%; Score 29; DB 10; Length 273;
 Best Local Similarity 80.0%; Pred. No. 3.1e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FPMWR 5
 Db 79 FPMWR 83
 RESULT 9
 O9HT83 PRELIMINARY: PRT: 292 AA.
 AC 09HT83;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

IE HYPOTHETICAL PROTEIN PA5488.
 CN PA5488.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 CC Pseudomonas
 CX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Mong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Saiter M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an
 opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 DR EMBL: AE004961; AAC08873.1; -;
 DR InterPro: IPR000504; RRM.
 DR PROSITE: PS00030; RRM_RNP_1; UNKNOWN_1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 292 AA; 32949 MW; 1461037E8CCDFD4 CRC64;

Query Match: 96.7%; Score 29; DB 2; Length 292;
 Best Local Similarity 80.0%; Pred. No. 3.3e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 FPMWR 5
 Db 231 FPMWR 235

RESULT 10
 ID 09D6L7 PRELIMINARY: PRT: 321 AA.
 AC 09D6L7;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE 2310076014RIK PROTEIN.
 DE 2310076014RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TONGUE;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arikawa T., Hara A., Fukunishi Y., Komio H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasuwa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Balalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schirral L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL: AK010208; BAB26769.1; -;

DR MCD:1919189: 2310076014Rik.
 DR InterPro: IPR000847; HTH_LYSR.
 DR PROSITE: PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.
 SO SEQUENCE 321 AA: 35038 MW: 66275BC746552DFB CRC64;

Query Match 96.7%; Score 29; DB 11; Length 321;
 Best Local Similarity 80.0%; Pred. No. 3.7e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FPXWR 5
 Db 44 FPSWR 48

RESULT 11
 ID 09APJ5 PRELIMINARY: PRT: 332 AA.
 AC 09APJ5;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE ENDO-1,4-BETA-D-GLUCANASE PRECURSOR.
 CN CEL8Y.
 OS Erwiniia chrysanthemi.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Pectobacterium.
 OX NCBI_TaxID=556;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=PY35;
 RA Cho S.J., Park S.R., Yun H.D.;
 RT "Cloning and sequencing of a cel8Y gene of Pectobacterium chrysanthemi
 RT PY35.";
 RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF282321; AAG49556.1; -.
 KW Signal.
 FT SIGNAL. 1 23 POTENTIAL.
 SO SEQUENCE 332 AA: 37627 MW: A1B5D7B0CB20EE6 CRC64;

Query Match 96.7%; Score 29; DB 2; Length 332;
 Best Local Similarity 80.0%; Pred. No. 3.8e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FPXWR 5
 Db 178 FPSWR 182

RESULT 12
 ID 09L3G9 PRELIMINARY: PRT: 333 AA.
 AC 09L3G9;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE ENOGLUCANASE PRECURSOR (EC 3.2.1.4).
 CN CELA.
 OS Erwiniia rhapontici.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Erwiniia.
 OX NCBI_TaxID=55212;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCPPB2989;
 RA Saarialhti H.T.;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN 121
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCPPB2989;
 RA Rieki R.;
 RT "Members of the amylovora group of Erwiniia are cellulolytic and

KT possess genes homologous to the type II secretion pathway.";
 RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AJ276358; CAB89803.1; -.
 DR InterPro: IPR002037; Glyco_hydro_8.
 CR Pfam: PF01270; Glyco_hydro_8; 1.
 DR PRINTS: PR00735; GLYHDLASE8.
 KW Signal; Hydrolase; Glycosidase.
 FT SIGNAL. 1 23 POTENTIAL.
 FT CHAIN 24 333 ENDOGLUCANASE.
 SO SEQUENCE 333 AA: 37783 MW: D75CFE212302673A CRC64;

Query Match 96.7%; Score 29; DB 2; Length 333;
 Best Local Similarity 80.0%; Pred. No. 3.8e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FPXWR 5
 Db 178 FPSWR 182

RESULT 13
 ID 09FTM4 PRELIMINARY: PRT: 334 AA.
 AC 09FTM4;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-JUN-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE P0005A05.25 PROTEIN.
 CN P0005A05.25.
 OS Oryza sativa (Rice).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 CC Ehrhartoideae; Oryzae; Oryza.
 OX NCBI_TaxID=4530;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 RT clone:P0005A05.";
 RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AP002863; BAB16921.1; -.
 DR InterPro: IPR000520; Exonuclease.
 DR InterPro: IPR000822; Znf-C2H2.
 CR Pfam: PF00929; Exonuclease; 1.
 DR Pfam: PF00096; ZF-C2H2; 1.
 CR SMART: SM00479; EXO111; 1.
 CR SMART: SM00355; Znf_C2H2; 2.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
 DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 1.
 KW DNA-binding; Zinc-finger.
 SO SEQUENCE 334 AA: 37172 MW: A34531D625704A88 CRC64;

Query Match 96.7%; Score 29; DB 10; Length 334;
 Best Local Similarity 80.0%; Pred. No. 3.8e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FPXWR 5
 Db 301 FPSWR 305

RESULT 14
 ID 015429 PRELIMINARY: PRT: 349 AA.
 AC 015429;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE CD46.
 OS Homo sapiens (Human).

OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
 CC Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.
 OX NCBI_TaxID=9606;

RA TISSUE=TESTIS;
 RP SEQUENCE FROM N.A.

RT "cDNA cloning and characterization of human sperm CD46."
 RL Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.

DR EMBL: D84105; BAA12224.1; -

DR HSSP: P10998; 1VVD.

DR InterPro: IPR000436; Sushl_SCR_CCP.

DR Pfam: PF00084; sushl: 4.

DR SMART: SM00032; CCP: 4.

SO SEQUENCE 349 AA; 39325 MW; 8EFCEDA30D3C818E CRC64;

Query Match 96.7%; Score 29; DB 4; Length 349;
 Best Local Similarity 80.0%; Pred. No. 4e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 DB 11 FPXMR 15

RESULT 15

O9SAA1 PRELIMINARY: PRT: 417 AA.

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)

DE F25C20.9.

GN F25C20.9.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:

OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae:

OC eurosids 11; Brassicales; Brassicaceae; Arabidopsids.

OX NCBI_TaxID=3702;

RA SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA:

RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,

Li J., Lee J.M., Kremetskaia I., Luros J., Ngan I., Liu A.,

Gonzalez A., Altafi H., Araujo R., Chao Q., Conn L., Conway A.B.,

Dunn P., Hansen N., Hultzer L., Kim C., Palm C., Rowley D., Shin P.,

Walker M., Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.,

RT Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.

RL Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA:

RA Theologis A.;

RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA:

RA Theologis;

RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.

DR EMBL: AC007296; AAD30247.1; -

SO SEQUENCE 417 AA; 45985 MW; B86EBF06AC19A63D CRC64;

Query Match 96.7%; Score 29; DB 10; Length 417;
 Best Local Similarity 80.0%; Pred. No. 4.7e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 DB 377 FPXMR 381

Search completed: February 27, 2002, 11:54:08
 Job time: 228 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:50:37 : Search time 32.48 Seconds
(without alignments)
3,464 Million cell updates/sec

Title: US-09-446-109a-22

Perfect score: 30

Sequence: 1 FPXMR 5

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfilest.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	96.7	38	1 US-08-176-500-118	Sequence 118, App
2	29	96.7	38	1 US-08-471-052A-118	Sequence 118, App
3	29	96.7	38	1 US-08-189-331-118	Sequence 118, App
4	29	96.7	38	2 US-08-471-939-118	Sequence 118, App
5	29	96.7	38	2 US-08-471-800-118	Sequence 118, App
6	29	96.7	38	2 US-08-471-068-118	Sequence 118, App
7	29	96.7	241	4 US-08-823-120-1	Sequence 1, Appl
8	29	96.7	324	2 US-08-528-057-46	Sequence 46, Appl
9	29	96.7	370	2 US-08-528-057-42	Sequence 42, Appl
10	29	96.7	372	1 US-08-202-056-5	Sequence 5, Appl
11	29	96.7	372	1 US-08-076-093A-6	Sequence 6, Appl
12	29	96.7	372	1 US-08-701-265-6	Sequence 6, Appl
13	29	96.7	372	2 US-08-284-586-6	Sequence 6, Appl
14	29	96.7	372	2 US-08-805-478-6	Sequence 6, Appl
15	29	96.7	372	2 US-08-807-477A-6	Sequence 6, Appl
16	29	96.7	372	2 US-08-801-238-6	Sequence 6, Appl
17	29	96.7	372	2 US-08-801-228-6	Sequence 6, Appl
18	29	96.7	372	2 US-09-104-296-6	Sequence 6, Appl
19	29	96.7	372	3 US-08-982-493-8	Sequence 8, Appl
20	29	96.7	373	3 US-08-528-057-44	Sequence 44, Appl
21	29	96.7	377	2 US-08-528-057-2	Sequence 2, Appl
22	29	96.7	384	4 US-08-139-195-2	Sequence 2, Appl
23	29	96.7	384	6 US-08-139-195-2	Sequence 2, Appl
24	28	93.3	57	1 US-08-370-225-29	Sequence 29, Appl
25	28	93.3	57	1 US-08-370-225-30	Sequence 29, Appl
26	28	93.3	57	1 US-08-461-859-29	Sequence 29, Appl
27	28	93.3	57	1 US-08-461-859-30	Sequence 30, Appl

28	28	93.3	57	5 PCT-US93-10069-29	Sequence 29, Appl
29	28	93.3	57	5 PCT-US93-10069-30	Sequence 30, Appl
30	28	93.3	298	2 US-08-061-636-3	Sequence 3, Appl
31	28	93.3	298	2 US-08-874-347-19	Sequence 19, Appl
32	28	93.3	298	3 US-09-093-522-19	Sequence 19, Appl
33	28	93.3	298	5 PCT-US94-05266-3	Sequence 3, Appl
34	28	93.3	373	2 US-08-846-762-13	Sequence 13, Appl
35	28	93.3	374	3 US-08-982-493-6	Sequence 6, Appl
36	28	93.3	437	2 US-09-031-059-1	Sequence 1, Appl
37	28	93.3	437	2 US-09-031-059-3	Sequence 3, Appl
38	28	93.3	760	1 US-08-195-152-2	Sequence 2, Appl
39	26	86.7	10	4 US-09-461-697-409	Sequence 409, App
40	26	86.7	12	4 US-09-461-697-407	Sequence 407, App
41	26	86.7	113	2 US-07-668-648-10	Sequence 10, Appl
42	26	86.7	113	2 US-08-429-898-10	Sequence 10, Appl
43	26	86.7	113	2 US-08-431-333-10	Sequence 10, Appl
44	26	86.7	113	5 PCT-US91-02321-10	Sequence 10, Appl
45	26	86.7	862	1 US-08-325-267A-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-176-500-118
Sequence 118, Application US/08176500
Patent No. 5498538
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
ATTORNEY/AGENT INFORMATION:
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 141
CORRESPONDENCE ADDRESS:
ADDRESS: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/176,500
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/013,416
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-143
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-176-500-118
Query Match 96.7% Score 29; DB 1; Length 38;
Best Local Similarity 80.0% Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FPXMR 5
DB 10 FPSMR 14

RESULT 2

US-08-471-052A-118
Sequence 118, Application US/08471052A
Patent No. 3625033
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
APPLICANT: Fowles, D. M.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 166
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471.052A
FILING DATE: 06-JUNE-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-179
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-471-052A-118

Query Match 96.7%; Score 29; DB 1; Length 38;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
DB 10 FPSMR 14

RESULT 3

US-08-189-331-118
Sequence 118, Application US/08189331
Patent No. 5747334
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
APPLICANT: Fowles, D. M.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 186
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/189.331
FILING DATE: Concurrently herewith
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-155
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-189-331-118

Query Match 96.7%; Score 29; DB 1; Length 38;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
DB 10 FPSMR 14

RESULT 4

US-08-471-939-118
Sequence 118, Application US/08471939
Patent No. 5844076
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
APPLICANT: Fowles, D. M.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 141
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471.939
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/013,416
FILING DATE: 01-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-143
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:

SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-471-939-118

Query Match 96.7%; Score 29; DB 2; Length 38;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
1111
DB 10 FPSMR 14

RESULT 5
US-08-471-800-118
Sequence 118, Application US/08471800
Patent No. 5852167
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
APPLICANT: Fowlkes, D. M.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 141
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennile & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,800
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/013,416
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-143
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-471-800-118

Query Match 96.7%; Score 29; DB 2; Length 38;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
1111
DB 10 FPSMR 14

RESULT 6
US-08-471-068-118
Sequence 118, Application US/08471068
Patent No. 5948635
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
APPLICANT: Fowlkes, D. M.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 186
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennile & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,068
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/189,331
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-155
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-471-068-118

Query Match 96.7%; Score 29; DB 2; Length 38;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
1111
DB 10 FPSMR 14

RESULT 7
US-08-823-120-1
Sequence 1, Application US/08823120
Patent No. 6149919
GENERAL INFORMATION:
APPLICANT: Domenighini, Mario
APPLICANT: Rappuoli, Rino
APPLICANT: Pizza, Mariagrazia
TITLE OF INVENTION: Immunogenic Detoxified Mutants of
TITLE OF INVENTION: Cholera Toxin and of the Toxin Lt, Their Preparation and
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chilton Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: California
COUNTRY: USA

ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/823,120
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/256,003
FILING DATE: 11-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0315,001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 601-2708
TELEFAX: (510) (655-3542
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 241 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-823-120-1

Query Match 96.7%; Score 29; DB 4; Length 241;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PPSMR 5
1111
Db 169 PPSMR 173

RESULT 8
US-08-528-057-46
Sequence 46, Application US/08528057
Patent No. 5846715
GENERAL INFORMATION:
APPLICANT: PURCELL, Damian F. J.
APPLICANT: RUSSELL, Sarah M.
APPLICANT: MCKENZIE, Ian F. C.
TITLE OF INVENTION: CD46 VARIANTS
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/528,057
FILING DATE: CONCURRENTLY HEREMITH
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/961,686
FILING DATE: 11-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU91/00199
FILING DATE: 10-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PK0133/90
FILING DATE: 11-MAY-1990

ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 17227/112 DACO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 324 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-528-057-46

Query Match 96.7%; Score 29; DB 2; Length 324;
Best Local Similarity 80.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PPSMR 5
1111
Db 11 PPSMR 15

RESULT 9
US-08-528-057-42
Sequence 42, Application US/08528057
Patent No. 5846715
GENERAL INFORMATION:
APPLICANT: PURCELL, Damian F. J.
APPLICANT: RUSSELL, Sarah M.
APPLICANT: MCKENZIE, Ian F. C.
TITLE OF INVENTION: CD46 VARIANTS
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/528,057
FILING DATE: CONCURRENTLY HEREMITH
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/961,686
FILING DATE: 11-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU91/00199
FILING DATE: 10-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PK0133/90
FILING DATE: 11-MAY-1990
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 17227/112 DACO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 370 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-528-057-42

Query Match 96.7%; Score 29; DB 1; Length 370;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPMR 5
11 11
DB 11 FPMR 15

RESULT 10
US-08-202-056-5
Sequence 5, Application US/08202056
Patent No. 5440021
GENERAL INFORMATION:
APPLICANT: Chuntharapai, Anan
APPLICANT: Hebert, Caroline
APPLICANT: Kim, Kyung Jin
APPLICANT: Lee, James
TITLE OF INVENTION: Antibodies to Human IL-8 Type B Receptor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 KB floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202.056
FILING DATE: 25-FEB-1994
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/677211
FILING DATE: 29-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B.
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: 706P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-202-056-5

Query Match 96.7%; Score 29; DB 1; Length 372;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPMR 5
11 11
DB 352 FPMR 356

RESULT 11
US-08-076-093A-6
Sequence 6, Application US/08076093A
Patent No. 5543503
GENERAL INFORMATION:

APPLICANT: Chuntharapai, Anan
APPLICANT: Lee, James
APPLICANT: Hebert, Caroline
APPLICANT: Jin Kim, K.
TITLE OF INVENTION: Antibodies to Human PFA Receptors
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 MB floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Minipatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/076.093A
FILING DATE: 11-Jun-1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/677211
FILING DATE: 29-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B.
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: 706P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: Amino acid
TOPOLOGY: linear
US-08-076-093A-6

Query Match 96.7%; Score 29; DB 1; Length 372;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPMR 5
11 11
DB 352 FPMR 356

RESULT 12
US-08-701-265-6
Sequence 6, Application US/08701265
Patent No. 5776457
GENERAL INFORMATION:
APPLICANT: Chuntharapai, Anan
APPLICANT: Lee, James
APPLICANT: Hebert, Caroline
APPLICANT: Jin Kim, K.
TITLE OF INVENTION: Antibodies to Human PFA Receptors
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 MB floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/701.265
FILING DATE: 22-AUG-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/076093
FILING DATE: 11-JUN-1993
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/677211
FILING DATE: 29-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: 706P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-701-265-6

Query Match 96.7% Score 29; DB 1; Length 372;
Best Local Similarity 80.0% Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
1111
DB 352 FPSMR 356

RESULT 13
US-08-284-586-6
Sequence 6, Application US/08284586
Patent No. 5840856
GENERAL INFORMATION:
APPLICANT: Chuntharapal, Anan
APPLICANT: Lee, James
APPLICANT: Hebert, Caroline
APPLICANT: Jin Kim, K
TITLE OF INVENTION: Antibodies to Human PFAA Receptors
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 MB floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/284,586
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: 706P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-284-586-6

APPLICATION NUMBER: 07/677211
FILING DATE: 29-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: 706P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-284-586-6

Query Match 96.7% Score 29; DB 2; Length 372;
Best Local Similarity 80.0% Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
1111
DB 352 FPSMR 356

RESULT 14
US-08-805-478-6
Sequence 6, Application US/08805478
Patent No. 5874543
GENERAL INFORMATION:
APPLICANT: Chuntharapal, Anan
APPLICANT: Lee, James
APPLICANT: Hebert, Caroline
APPLICANT: Jin Kim, K
TITLE OF INVENTION: ANTIBODIES TO PFAA RECEPTOR
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 MB floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/805,478
FILING DATE: 25-Feb-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/284586
FILING DATE: 10-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/076093
FILING DATE: 11-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: 70706P2P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:

LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-805-478-6

Search completed: February 27, 2002, 11:50:37
Job time: 866 sec

Query Match 96.7%: Score 29; DB 2; Length 372;
Best Local Similarity 80.0%: Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
Db 352 FPSWR 356

RESULT 15
US-08-802-627A-6
Sequence 6, Application US/08802627A
Patent No. 5892017
GENERAL INFORMATION:
APPLICANT: Lee, James
APPLICANT: Wood, William I.
TITLE OF INVENTION: NUCLEIC ACID ENCODING PFAA RECEPTOR
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 MB floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatlin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/802,627A
FILING DATE: 19-Feb-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/284586
FILING DATE: 10-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/076093
FILING DATE: 11-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B.
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: P0706P2P1D2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO.: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-802-627A-6

Query Match 96.7%: Score 29; DB 2; Length 372;
Best Local Similarity 80.0%: Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
Db 352 FPSWR 356

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:51:50 : Search time 66.26 Seconds
(without alignments)
5,590 Million cell updates/sec

Title: US-09-446-109A-23
Sequence: 1 FPXWR 5

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SID58/gcgdata/geneseq/geneseqp/AA1980.DAT.*
2: /SID58/gcgdata/geneseq/geneseqp/AA1981.DAT.*
3: /SID58/gcgdata/geneseq/geneseqp/AA1982.DAT.*
4: /SID58/gcgdata/geneseq/geneseqp/AA1983.DAT.*
5: /SID58/gcgdata/geneseq/geneseqp/AA1984.DAT.*
6: /SID58/gcgdata/geneseq/geneseqp/AA1985.DAT.*
7: /SID58/gcgdata/geneseq/geneseqp/AA1986.DAT.*
8: /SID58/gcgdata/geneseq/geneseqp/AA1987.DAT.*
9: /SID58/gcgdata/geneseq/geneseqp/AA1988.DAT.*
10: /SID58/gcgdata/geneseq/geneseqp/AA1989.DAT.*
11: /SID58/gcgdata/geneseq/geneseqp/AA1990.DAT.*
12: /SID58/gcgdata/geneseq/geneseqp/AA1991.DAT.*
13: /SID58/gcgdata/geneseq/geneseqp/AA1992.DAT.*
14: /SID58/gcgdata/geneseq/geneseqp/AA1993.DAT.*
15: /SID58/gcgdata/geneseq/geneseqp/AA1994.DAT.*
16: /SID58/gcgdata/geneseq/geneseqp/AA1995.DAT.*
17: /SID58/gcgdata/geneseq/geneseqp/AA1996.DAT.*
18: /SID58/gcgdata/geneseq/geneseqp/AA1997.DAT.*
19: /SID58/gcgdata/geneseq/geneseqp/AA1998.DAT.*
20: /SID58/gcgdata/geneseq/geneseqp/AA1999.DAT.*
21: /SID58/gcgdata/geneseq/geneseqp/AA2000.DAT.*
22: /SID58/gcgdata/geneseq/geneseqp/AA2001.DAT.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	96.7	10	AAV08336	O. hannah venom al
2	29	96.7	11	AAV08334	O. hannah venom al
3	29	96.7	38	AAV08491	Binding domain of
4	29	96.7	94	AAV06599	Rat sodium channel
5	29	96.7	94	AAV20130	Rat truncated sodi
6	29	96.7	135	AAV04918	Myobacterium spec
7	29	96.7	177	AAV58892	Breast and ovarian
8	29	96.7	299	AAV52488	Helicobacter pylori
9	29	96.7	314	AAV93943	CD46 construct del
10	29	96.7	324	AAV15233	CD46 from clone pm
11	29	96.7	359	AAV0396	Cellulase. Cellul

12	29	96.7	372	13	AAV27793	New platelet facto
13	29	96.7	372	16	AAV92239	Chemokine superfam
14	29	96.7	372	16	AAV68813	Human lymphocyte p
15	29	96.7	372	20	AAV06644	Human Burkitt's ly
16	29	96.7	372	21	AAV90627	Human G protein-co
17	29	96.7	372	21	AAV90661	Human mutant G pro
18	29	96.7	373	12	AAV15230	CD46 from clone pm
19	29	96.7	376	17	AAV93939	CD46 wild-type. H
20	29	96.7	377	12	AAV15229	CD46 from clone pm
21	29	96.7	377	17	AAV93942	CD46 construct sub
22	29	96.7	377	17	AAV93941	CD46 construct sub
23	29	96.7	384	18	AAV27484	Human MCP. Homo s
24	29	96.7	384	12	AAV10924	Human membrane cof
25	29	96.7	384	16	AAV86316	Human CD46. Homo
26	29	96.7	418	21	AAV13589	Streptomyces globi
27	29	96.7	421	21	AAV85834	Lung cancer associ
28	29	96.7	421	22	AAV75528	Human colon cancer
29	29	96.7	497	20	AAV04972	Myobacterium spec
30	29	96.7	520	22	AAV90739	C glutamicum prote
31	29	96.7	589	20	AAV34524	Porphyromonas ging
32	29	96.7	590	21	AAV94928	Human secreted pr
33	29	96.7	593	20	AAV34523	Porphyromonas ging
34	29	96.7	606	20	AAV34393	Porphyromonas ging
35	29	96.7	611	22	AAV03762	CAB-2 chimeric pro
36	29	96.7	612	22	AAV84488	Human protein sequ
37	29	96.7	614	18	AAV20991	H. pylori inner me
38	29	96.7	621	22	AAV84049	Human protein sequ
39	29	96.7	634	22	AAV40125	Human polypeptide
40	29	96.7	634	22	AAV92742	Human protein sequ
41	29	96.7	634	22	AAV83478	Human protein sequ
42	29	96.7	634	22	AAV85120	Human protein sequ
43	29	96.7	1464	19	AAV9294	An antigen from de
44	29	96.7	42	22	AAV15108	Peptide #1542 enco
45	29	93.3	42	22	AAV27563	Peptide #1600 enco

ALIGNMENTS

RESULT 1	AAV08336	standard; Peptide: 10 AA.
XX	AAV08336:	
NC	19-JUL-1999 (first entry)	
DT	O. hannah venom alpha-neurotoxin homologous peptide 5.	
XX		
DE	O. hannah venom alpha-neurotoxin homologous peptide 5.	
XX		
RW	Snake; venom. King cobra; alpha-neurotoxin; toxin; analgesia; screening.	
XX		
GS	Ophiophagus hannah.	
XX		
PN	W09924055-A1.	
XX	20-MAY-1990	
TD		
XX	03-NOV-1998; 98WO-SG00087.	
PF		
XX	06-NOV-1997; 97SG-0003972.	
TR		
XX	(UYSI-) UNTV SINGAPORE NAT.	
FA		
XX	Copaladishnakone P, Gwee MCE, Kint RM, Pu XC, Wong PT;	
XX	WPI: 1999-327205/27.	
DR		
XX	Snake neurotoxin derived peptides	
XX		
PT	Snake neurotoxin derived peptides	
XX		
FS	Claim 11; Page 43; 46pp; English.	
XX	This invention describes novel peptide derivatives of alpha-neurotoxin,	
CC	derived from snake venom toxins of Ophiophagus hannah and which are	

CC capable of inducing or facilitating analgesia. The peptides, optionally
 CC with a carrier and/or diluent, are used to induce analgesia in a subject.
 CC Antibodies derived from the products of the invention can be used to
 CC screen for subject peptides, e.g. for purification purposes.
 XX
 SQ Sequence 10 AA:

Query Match 96.7%; Score 29; DB 20; Length 10;
 Best Local Similarity 80.0%; Pred. No. 13;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR-5
 1111
 DB 3 fplwr 7

RESULT 2

AAV08334
 ID AAV08334 standard; Peptide: 11 AA.

XX
 AC AAV08334;

XX 19-JUL-1999 (first entry)

DE O. hannah venom alpha-neurotoxin homologous peptide 3.

XX Snake; venom: King cobra; alpha-neurotoxin; toxin; analgesia; screening.

XX Ophiophagus hannah.

XX WO9924055-A1.

XX 20-MAY-1999.

XX 03-NOV-1998; 98WO-SG00087.

XX 06-NOV-1997; 97SG-0003972.

XX (UYSI-) UNIV SINGAPORE NAT.

XX Copladrishnakone P, Gwee MCE, Klini RM, Pu XC, Wong PT;

DR WPI: 1999-327205/27.

XX Snake neurotoxin derived peptides

XX Claim 9; Page 42; 46pp; English.

CC This invention describes novel peptide derivatives of alpha-neurotoxin,
 CC derived from snake venom toxins of Ophiophagus hannah and which are
 CC capable of inducing or facilitating analgesia. The peptides, optionally
 CC with a carrier and/or diluent, are used to induce analgesia in a subject.
 CC Antibodies derived from the products of the invention can be used to
 CC screen for subject peptides, e.g. for purification purposes.
 XX
 SQ Sequence 11 AA:

Query Match 96.7%; Score 29; DB 20; Length 11;
 Best Local Similarity 80.0%; Pred. No. 14;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 1111
 DB 3 fplwr 7

RESULT 3

AAV08491
 ID AAV08491 standard; Protein: 38 AA.

XX
 AC AAV08491;

XX 18-APR-1995 (first entry)

XX Binding domain of a polystyrene-binding TSAR (SB-9-5).

XX TSAR: totally synthetic affinity reagent; synthetic; binding domain;
 KW effector domain; concatenated heterofunctional protein; linker;
 KW direct; rapid; detection; screening; treatment; monoclonal antibody;
 KW Sm antigen; mouse model; autoimmune disease; SLE; MAb;
 KW systemic lupus erythematosus; polystyrene; SB.

XX Synthetic.

XX WO9418318-A.

XX 18-AUG-1994.

XX 01-FEB-1994; 94WO-US00977.

XX 01-FEB-1993; 93US-0013416.

XX 30-DEC-1993; 93US-0176500.

XX 31-JAN-1994; 94US-0189331.

XX (UYN-) UNIV NORTH CAROLINA.

XX Fowlkes DM, Kay BK;

XX WPI: 1994-279739/34.

XX Identifying proteins or peptide(s) which bind a ligand - by
 PT screening a recombinant vector library expressing fusion proteins
 PT comprising a binding domain and an effector domain

XX Example 7.8; Page 120; 255pp; English.

CC AAV08491-507 are amino acid sequences of the binding domain of TSAR
 CC (Totally Synthetic Affinity Reagents) peptides. These particular
 CC peptides are encoded by the polystyrene (SR) binding phage. There
 CC is no apparent linear motif, but the peptides are rich in Trp, Tyr and
 CC Gly, poor in Arg, Val and Lys, and completely lack cysteine residues.
 CC The non-variable amino acids at the NH2 and COOH terminals are
 CC not shown. TSAR peptides are generated using generic oligonucleotides
 CC (see AAO70470-73 for examples). TSARs are concatenated heterofunctional
 CC proteins or peptides, comprising at least two functional regions - a
 CC binding domain with affinity for a ligand and a second effector peptide
 CC portion that is chemically or biologically active. They may further
 CC comprise a linker peptide between the 2 domains. The TSARs or compns.
 CC comprising a TSAR binding domain can be used in vivo to deliver a
 CC chemically or biologically active moiety, eg. metal ion, radioisotope,
 CC peptide, toxin or enzyme, to the specific target or on the cell. They
 CC can also replace the function of macromolecules eg. monoclonal or
 CC polyclonal antibodies and therefore circumvent the need for complex
 CC methods of hybridoma formation or in vivo antibody production. The TSARs
 CC are easily characterised and have designed activity allowing direct and
 CC rapid detection in a screening process.
 XX
 SQ Sequence 38 AA:

Query Match 96.7%; Score 29; DB 15; Length 38;
 Best Local Similarity 80.0%; Pred. No. 46;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 1111
 DB 10 fplwr 14

RESULT 4

AAV06599
 ID AAV06599 standard; Protein: 94 AA.

XX
 AC AAV06599;

XX	26-OCT-1999	(first entry)	
DT			
XX			
DE	Rat sodium channel Nan splice variant C-terminal sequence.		
XX			
KW	Nan; sodium channel; ion transport; rat; dorsal root ganglia;		
KW	pain; paraesthesia; hyperexcitability; therapy.		
XX			
OS	Rattus sp.		
XX			
XX	MO9936889-A2.		
XX			
PD	05-AUG-1999.		
XX			
PF	29-JAN-1999; 99WO-US02008.		
XX			
PR	20-NOV-1998; 98US-0109402.		
XX			
PR	29-JAN-1998; 98US-0072990.		
XX			
PA	(UYVA) UNITV YALE.		
XX			
PI	D1b-HaJj S, Waxman S;		
XX			
DR	WPI: 1999-479168/40.		
XX			
PT	New isolated nucleic acids encoding sodium channels, used to develop		
PT	products for treating acute or chronic pain or hyperexcitability		
PT	phenomena		
XX			
PS	Example 5; Page 35; 91pp; English.		
XX			
CC	This is the C-terminal sequence of a splice variant of rat sodium		
CC	channel Nan. This truncated variant comprises amino acids 1-1378		
CC	of the Nan sequence given in AAY06596 and the present, 94-amino acid		
CC	sequence at the C-terminal end. It thus lacks the 387 C-terminal		
CC	amino acids of full-length Nan. The variant arises from the use		
CC	of a cryptic donor splice site in exon 23 and a novel exon 23'		
CC	which is located in intron 23 of the rat Nan gene. Nan is a		
CC	previously unidentified voltage gated sodium channel protein that		
CC	is preferentially expressed in dorsal root ganglia or trigeminal		
CC	ganglia, and which produces a TTX-R sodium current. The invention		
CC	provides methods for identifying agents that modulate Nan channel		
CC	activity or expression and for using such agents to treat acute or		
CC	chronic pain, paraesthesia and hyperexcitability phenomena.		
XX			
SQ	Sequence 94 AA:		
	Query Match 96.7%; Score 29; DB 20; Length 94;		
	Best Local Similarity 80.0%; Pred. No. 1.1e+02;		
	Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
OY	1 FPXMR 5		
	1111		
DB	52 fpawr 56		
RESULT 5			
AAB20130			
ID AAB20130	standard; Protein; 94 AA.		
XX			
AC	AAB20130;		
XX			
DT	30-APR-2001 (first entry)		
XX			
DE	Rat truncated sodium channel C-terminal region.		
XX			
KW	Sodium channel; Nan; rat; tetrodotoxin resistant; pain;		
KW	paraesthesia; hyperexcitability; analgesic; vaccine; therapy;		
KW	diagnosis.		
XX			
OS	Rattus norvegicus.		
XX			

FN WO200105831-A1.
 XX
 PD 25-JAN-2001.
 XX
 PF 14-JUL-2000; 2000WO-US19342.
 XX
 PR 16-JUL-1999; 99US-0354147.
 XX
 PA (UYRA) UNIV YALE.
 XX
 PI DiB-Hajj S, Waxman SG;
 ER
 XX
 FI MPI: 2001-103147/11.
 XX
 PT Nucleic acid molecules encoding human tetrodotoxin resistant sodium
 XX channels, useful for preventing, diagnosing and treating pain,
 XX parasthesia and/or hyperexcitability phenomena -
 XX
 PS Example 5; Page 34; 162pp; English.
 XX
 CC The present sequence is that of the C-terminal portion of a
 CC truncated rat NaN variant. The variant lacks the 387 C-terminal
 CC amino acids of full-length rat NaN (see AAB20122), which are
 CC replaced by the present 94-amino acid polypeptide. The N-terminal
 CC 1378 amino acids of the truncated variant are identical to those
 CC of full-length NaN. The new sequence arises from the use of a
 CC cryptic donor splice site in exon 23 and a novel exon 23' which is
 CC located in intron 23. Rat NaN is a novel tetrodotoxin resistant
 CC sodium channel NaN belonging to the α -subunit voltage-gated sodium
 CC channel protein family. These sodium channels underlie the
 CC generation and propagation of impulses in excitable cells such as
 CC neurons and muscle fibres. Preferential expression of NaN on
 CC sensory neurons innervating the body (dorsal root ganglia) and
 CC the face (trigeminal ganglia), but not on other neurons, makes
 CC it a very useful target for diagnostic and/or therapeutic uses in
 CC relation to acute and/or chronic pain pathologies. A claimed
 CC method of treating pain, parasthesia and/or hyperexcitability
 CC phenomena in a human or animal subject involves administering an
 CC agent that alters sodium current flow through NaN channels, or
 CC which modulates transcription or translation of NaN mRNA, in
 CC dorsal root ganglia or trigeminal neurons. NaN polypeptides can
 CC be obtained by recombinant expression, and used to treat disorders
 CC associated with decreased sodium channel expression, to screen for
 CC compounds that modulate sodium channel expression or activity,
 CC and to raise antibodies useful as diagnostic agents.
 CC
 XX
 SQ Sequence 94 AA:
 XX
 QY 1 FPXMR 5
 XX 1111
 DB 52 fpawr 56
 XX
 RESULT 6
 AA04918
 ID AA04918 standard; Protein: 135 AA.
 XX
 AC AA04918;
 XX
 DT 06-JUL-1999 (first entry)
 XX
 DE Mycobacterium species protein sequence 35B.
 XX
 KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
 XX hybridisation; detection; vaccine; immunisation; infection.
 XX
 OS Mycobacterium sp.
 XX

PN W09909186-A2.
 XX 25-FEB-1999.
 PD 14-AUG-1998; 98WO-FR01813.
 XX 11-SEP-1997; 97FR-0011325.
 PR 14-AUG-1997; 97FR-0010404.
 XX (INSP) INST PASTEUR.
 PA Gicquel B, Lim EM, Pellicic V, Portnoi D, Goguet de la Salmoniere Y;
 PI Guigueno A.
 DR WPI: 1999-181045/15.
 DR N-PSDB: AAX34171.
 XX Mycobacterial DNA vectors containing reporter constructs - for
 PT identifying coding or promoter sequences involved in
 PT infection-associated protein expression
 PS Claim 32: Fig 35B: 309pp: French.
 XX Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted
 CC proteins from various Mycobacterium species microorganisms. The
 CC encoding nucleotide sequences can be used as primers and probes for
 CC methods for detecting and identifying mycobacteria, especially belonging
 CC to the M. tuberculosis complex. The encoded proteins can be used in
 CC vaccines for immunisation against a bacterial or viral infection.
 XX Sequence 135 AA:
 SQ

Query Match 96.7%: Score 29; DB 20; Length 135;
 Best Local Similarity 80.0%: Pred. No. 1.5e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPMWR 5
 1111
 Db 66 fpmwr 70

RESULT 7
 AAB58892
 ID AAB58892 standard; Protein: 177 AA.
 XX AAB58892:

DT 27-MAR-2001 (first entry)

DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 600.

XX Human: breast cancer; cytostatic; immunosuppressive;
 KW neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 KW antidiabetic; antiinflammatory; antitumor; vulnereary; anticonvulsant;
 KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
 KW Addison's disease; allergy; autoimmune haemolytic anaemia;
 KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
 KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
 KW cardiovascular disorder; wound healing; neurological disease.
 XX Homo sapiens.
 OS
 XX W0200055173-A1.
 PN
 XX 21-SEP-2000.
 PD
 XX 08-MAR-2000; 2000WO-US05881.
 PF
 XX 12-MAR-1999; 99US-0124270.
 PR
 XX (IUMNA-) IUMMAN GENOME SCI INC.
 PA
 XX

PI Rosen CA, Ruben SM;
 XX WPI: 2000-611515/58.
 DR N-PSDB: AAF21795.
 XX New human breast and ovarian cancer associated gene sequences and the
 PT polypeptides encoded by these genes, useful in the prevention,
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular
 PT disorders and neurological diseases -
 XX Claim 11: Page 1037-1038; 1299pp: English.
 PS

XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
 CC associated with breast and ovarian cancer. Included in the invention are
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
 CC isolation and characterisation of the DNA and protein sequences of the
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist
 CC or antagonist sequences exhibit cytostatic; immunosuppressive;
 CC neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 CC antidiabetic; antiinflammatory; antitumor; vulnereary; anticonvulsant;
 CC antibacterial; antifungal; antiparasitic and cardiant activity. The
 CC polynucleotide and protein sequences are used in the diagnosis of cancer,
 CC particularly breast and ovarian cancer. The nucleic acid sequences,
 CC proteins, agonists and antagonists may also be used in the diagnosis,
 CC prevention and treatment of immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; cardiovascular disorders such as
 CC myocardial ischaemias; wound healing; neurological diseases such as
 CC cerebral anoxia and epilepsy; and infectious diseases.
 XX Sequence 177 AA:
 SQ

Query Match 96.7%: Score 29; DB 21; Length 177;
 Best Local Similarity 80.0%: Pred. No. 2e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPMWR 5
 1111
 Db 22 fpmwr 26

RESULT 8
 AAB52488
 ID AAB52488 standard; Protein: 299 AA.
 XX AAB52488:

DT 23-FEB-2001 (first entry)

DE Helicobacter pylori bait polypeptide #6.

XX Helicobacter pylori; two-hybrid system; protein-protein interaction;
 KW bait polypeptide; gastric ulcer; antibacterial.
 KW
 KW Helicobacter pylori.
 OS
 XX W0200066722-A1.
 PN
 XX 09-NOV-2000.
 PD
 XX 14-APR-2000; 2000WO-1B00603.
 PF
 XX 30-APR-1999; 99EP-0401066.
 PR
 XX (HYBR-) HYBRIGENICS SA.
 PA
 XX Legrain P, Selig L, Rain J;
 PI
 XX WPI: 2000-687535/67.
 DR N-PSDB: AAC97234.

XX A two-hybrid system for identifying compounds useful in the treatment
 of e.g. gastric ulcers comprises producing a collection of recombinant
 cell clones -

PS Example 5; Page 88-89; 267pp; English.

XX The present sequence is a bait polypeptide used in a Helicobacter
 pylori two-hybrid screen to identify protein-protein interactions.
 CC The method is used to identify a recombinant cell clone expressing a
 CC prey polypeptide which is capable of interacting with the bait
 CC polypeptide. The two hybrid system is useful for screening compounds
 CC for antibacterial activity. It may be used in the treatment of gastric
 CC ulcers. The polynucleotides are useful as amplification primers or
 CC specific detection probes. The polypeptides, vectors or host cells can
 CC be used as immunogens to produce mono- or polyclonal antibodies. The
 CC polynucleotides, polypeptides, antibodies, vectors, host cells or
 CC modulating agents can be used to produce a pharmaceutical composition.

XX Sequence 299 AA:

Query Match 96.7%; Score 29; DB 21; Length 299;
 Best Local Similarity 80.0%; Pred. No. 3.3e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 1111
 DB 89 fpswr 93

RESULT 9

AAR93943 ID AAR93943 standard; Protein: 314 AA.

XX AAR93943;

XX 21-MAY-1996 (first entry)

XX CD46 construct delSCR2/subSCR3.

XX CD46; recombinant protein; short consensus repeat; SCR;
 KW regulator of complement activation; transgenic animal; pig;
 KW organ transplantation.

XX Synthetic.

XX WO9606937-A1.

XX 07-MAR-1996.

XX 30-AUG-1995; 95WO-AU00553.

XX 30-AUG-1994; 94AU-0007724.

XX (AUST-) AUSTIN RES INST.

XX Christensen D, Loveland B, McKenzie JFC, Milland J;

XX WPI: 1996-160368/16.

XX N-PSDB: AAT17599.

XX Increasing prodn. of recombinant proteins, esp. CD46 - by reducing
 PT the amt. of A and/or T in an A and/or T rich region of encoding gene
 PT exon

XX Claim 12; Page 40-41; 60pp; English.

XX CD46 delSCR2/subSCR3 (AAR93943) is the product of a cDNA construct
 CC (AAT17599) obtd. by splice overlap extension PCR of wild-type CD46
 CC cDNA (AAT17595). The A-T content of A-T rich exon 5 of the gene,
 CC encoding the short consensus repeat (SCR) 3 of CD46, was lowered
 CC and SCR2 was deleted. This resulted in improved prodn. in

CC eukaryotic host cells, e.g. CHO-K1 and COS-7. CD46 delSCR2/subSCR3
 CC is used to prevent complement- or inflammation-mediated tissue
 CC damage, to improve immunity to tumours or viruses, to control
 CC fertilisation and to prevent spontaneous abortion. Expression in
 CC transgenic animals, esp. pigs, provides organs suitable for
 CC transplantation.

XX Sequence 314 AA:

Query Match 96.7%; Score 29; DB 17; Length 314;
 Best Local Similarity 80.0%; Pred. No. 3.4e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 1111
 DB 11 fpswr 15

RESULT 10

AAR15233 ID AAR15233 standard; Protein: 324 AA.

XX AAR15233;

XX 17-MAR-1992 (first entry)

XX CD46 from clone pm5.8.

XX SCR: short consensus repeats; transmembrane; cytoplasmic;
 KW membrane cofactor protein; MCP.

XX Homo sapiens.

XX OS

XX Key

XX Peptide 1..34 location/Qualifiers

XX Protein 34..324 /label= sig-peptide

XX Modified-site 83 /label= mat-protein

XX Modified-site 114 /label= N-glycosylation_site

XX Domain 289..304 /label= N-glycosylation_site

XX WO9118097-A

XX 28-NOV-1991

XX 10-MAY-1991; 91WO-AU00199.

XX 11-MAY-1990; 90AU-0000133.

XX (UYME-) UNIV MELBOURNE.

XX Purcell DJF, Russell SM, McKenzie JFC;

XX WPI: 1991-369251/50.

XX N-PSDB: AAQ14919.

XX New CD46 membrane co:factor protein variants - useful as probes
 PT to identify CD46 isoforms and for diagnosing spontaneous
 PT abortion, inhibiting immuno:response and treating leukaemia

XX Disclosure: fig 1A and 3B; 77pp; English.

XX The sequence of the pm5.8 clone is identical to the pm5.1 clone in
 CC the portion encoding the NH2 leader and four SCR regions. The
 CC sequence after nucleotide 890 was different, and results from
 CC reading through of the cDNA into an intron sequence after the
 CC fourth SCR. The protein encoded by this new sequence encodes
 CC a 16 amino acid hydrophobic region, probably serving as a membrane

CC spanning sequence. The putative cytoplasmic tail would also
 CC contain a high proportion of charged amino acids.
 CC See also AAU14915-25, AA015211-12 and AAR15457-59.
 XX
 SO Sequence 324 AA:

Query Match 96.7%; Score 29; DB 12; Length 324;
 Best Local Similarity 80.0%; Pred. No. 3.5e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 11 11
 DB 11 fpxmr 15

RESULT 11

AA070396
 ID AAP70396 standard; Protein: 359 AA.

XX
 AC AAP70396;

XX
 DT 30-APR-1991 (first entry)

XX
 DE Cellulase.

XX
 KW Cellulase; plasmid; transformation.

XX
 OS Cellulomonas uda CBA (FER.

XX
 PN JP62175178-A.

XX
 PD 31-JUL-1987.

XX
 PE 30-JAN-1986; 86JP-0016800.

XX
 PR 30-JAN-1986; 86JP-0016800.

XX
 PA (ENER-) SHIN ENERGY SOGO KAIHATSU KIKO.

XX
 PA (NENR-) NENRYOYO ALCOHOL KAIHATSU GIJUTSU KENKYU KUMIAI.

XX
 WP1: 1987-253827/36.

DR
 N-PSDB: AAN70617.

XX
 PT Bio-engineering cellulase prodn. - by preparing DNA chain,
 PT preparing plasmid contg. DNA chain in the gene expressible state,
 PT and transforming bacterial by plasmid

XX
 PS Disclosure: Fig 1(A-B); 12pp; Japanese.

XX
 CC Transformation E.coli JM103 (FERM P-8612) with the gene encoding
 CC this protein allows expression of the protein and easy and cheap
 CC prodn. of cellulase.

XX
 SO Sequence 359 AA:

Query Match 96.7%; Score 29; DB 8; Length 359;
 Best Local Similarity 80.0%; Pred. No. 3.9e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 11 11
 DB 178 fpxmr 182

RESULT 12

AA077793
 ID AAR27793 standard; Protein: 372 AA.

XX
 AC AAR27793;

XX
 DT 12-MAR-1993 (first entry)

XX
 DE New platelet factor 4 receptor superfamily member PF4ARL1.
 XX
 KW IL-8R; G-protein coupled receptor family; rhodopsin superfamily;
 KW pro-inflammatory cytokine; 8tr.9.

XX
 OS Homo sapiens.

XX
 PN M09217497-A.

XX
 PD 15-OCT-1992.

XX
 PE 23-MAR-1992; 92MO-US02317.

XX
 PR 29-MAR-1991; 91US-0677211.

XX
 PR 19-DEC-1991; 91US-0810782.

XX
 FA (GETH) GENENTECH INC.

XX
 PT Holmes WE, Lee J, Wood WI;

XX
 DR WP1: 1992-366191/44.

XX
 N-PSDB: AAQ37107.

XX
 PT Isolated human platelet factor 4 super-family receptor
 PT polypeptide and corresp. antibodies and DNA - useful as
 PT diagnostic and screening agents, and for treating inflammation or
 PT PF4AR-mediated disorders

XX
 PS Claim 7; Fig 5; 78pp; English.

XX
 CC The IL-8 receptor cDNA sequence was isolated (see AA029505) and a
 CC 874bp sub-fragment of the coding sequence was used as a probe to
 CC screen human cell line HL60 and human peripheral blood lymphocyte
 CC cDNA libraries. Two new gene sequences were found that are clearly
 CC related to the IL-8 receptor. One of these was contained in clone
 CC 8tr.9 and is predicted to encode an amino acid sequence which is
 CC 36% and 38% identical with the high and low affinity IL-8 receptor
 CC sequences, respectively. See also AAQ37107.

XX
 SO Sequence 372 AA:

Query Match 96.7%; Score 29; DB 13; Length 372;
 Best Local Similarity 80.0%; Pred. No. 4e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
 11 11
 DB 352 fpxmr 356

RESULT 13

AA092239
 ID AAR92239 standard; Protein: 372 AA.

XX
 AC AAR92239;

XX
 DT 26-MAR-1996 (first entry)

XX
 DE Chemokine superfamily receptor.

XX
 KW Interleukin; IL-8; inflammation; psoriasis; dermatitis;

XX
 KW Rheumatoid arthritis; inflammatory bowel disease;

XX
 KW Chronic lung inflammation; treatment; antibody;

XX
 KW affinity purification; detection.

XX
 OS Homo sapiens.

XX
 UN US5440021-A.

XX
 PD 08-AUG-1995.

PE 29-MAR-1991: 91US-0677211.
 XX 25-FEB-1994: 94US-0202056.
 PR 29-MAR-1991: 91US-0677211.
 XX (CHUN/) CHUNTHARAPAI A.
 PA (HEBE/) HEBERT C.
 PA (KIM/) KIM K J.
 PA (LEE/) LEE J.
 XX Chuntharapai A., Hebert C, Kim KJ, Lee J;
 XX WPI: 1995-283151/37.
 DR N-PSDB: AAQ99009.
 XX New antibodies against Interleukin 8 type B receptor - used to treat
 PT or prevent inflammation, also for detecting receptor expression and
 PT purification.
 XX Example 2: Columns 49-52: 62pp: English.
 PS Antibodies directed against the Interleukin-8 receptor B can be used
 CC to treat or prevent inflammation e.g. psoriasis, dermatitis.
 CC Rheumatoid arthritis and particularly inflammatory bowel disease and
 CC chronic lung inflammation. When immobilised, these antibodies may
 CC be used to detect Interleukin-8 receptor B expression in cells and
 CC tissues and for affinity purification of Interleukin-8 receptor B
 CC from cells. This sequence is an additional chemokine superfamily
 CC receptor which was identified by probing lambda libraries of genomic
 CC DNA from a human monocyte-like cell line (U-60) and human peripheral
 CC blood lymphocytes using a large fragment of the Interleukin-8 type
 CC A receptor DNA (See AAQ99006).
 XX Sequence 372 AA:

Query Match 96.7%; Score 29; DB 16; Length 372;
 Best Local Similarity 80.0%; Pred. No. 4e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
 ||||
 DB 352 fpswr 356

RESULT 14

AAR68813
 ID AAR68813 standard; Protein: 372 AA.

XX AAR68813;
 DT 18-JUL-1995 (first entry)

XX Human lymphocyte PPAR.

KW Interleukin-8 receptor; IL-8 receptor; PPAR;
 KW platelet factor superfamily receptor; lymphocyte; chemotactic;
 KW inflammation; inflammatory disease; arthritis; emphysema; cystic;
 KW fibrosis; colitis; bronchitis; meningitis; therapeutic.

XX Homo sapiens.

XX MO9428931-A.

XX 22-DEC-1994.

XX 07-JUN-1994: 94WO-US06380.

XX 11-JUN-1993: 93US-0076093.

XX (GETH) GENENTECH INC.

PA Chuntharapai A, Hebert C, Kim KJ, Lee J;
 PI

XX WPI: 1995-036114/05.
 DR N-PSDB: AAQ80522.
 XX Treatment of inflammatory disorders - by administering an
 PT antibody capable of binding a platelet factor 4 superfamily
 PT receptor polypeptide
 XX Disclosure: Page 56-58; 83pp: English.
 PS 2 PPAR members were identified by probing lambda libraries from
 CC human monocyte-like cell line HL-60 and human peripheral blood
 CC lymphocytes using a large fragment of IL-8 receptor DNA (full)
 CC sequence given in AAQ80520). The nucleotide sequences of the 2
 CC PPARs are given in AAQ80521 and AAQ80522, and their respective
 CC amino acid sequences in AAR68812 and AAR68813.
 XX Sequence 372 AA:

Query Match 96.7%; Score 29; DB 16; Length 372;
 Best Local Similarity 80.0%; Pred. No. 4e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

CY 1 FPXMR 5
 ||||
 DB 352 fpswr 356

RESULT 15

AAY06644
 ID AAY06644 standard; Protein: 372 AA.

XX AAY06644;

XX 26-OCT-1999 (first entry)

XX Human Burkitt's lymphoma receptor 1 (BLR1).

XX Burkitt's lymphoma receptor 1; BLR1; human;

XX B lymphocyte chemotactant; BLC; chemokine; ligand;

XX drug screening; leukaemia; autoimmune disease; therapy.

XX Homo sapiens.

XX WO9928468-A1.

XX 10-JUN-1999.

XX 02-DEC-1998: 98WO-US25561.

XX 02-DEC-1997: 97US-0982493.

XX (REGC) UNIV CALIFORNIA.

XX Cyster JG, Gunn MD, Williams LT;

XX WPI: 1999-493764/41.

XX N-PSDB: AAX87710.

XX Modulating interaction of a Burkitt's lymphoma Receptor 1

XX polypeptide and ligand, useful in drug screens

XX Claim 1: Page 38: 42pp: English.

CC This sequence represents human Burkitt's lymphoma receptor 1
 CC (BLR1). The invention relates to methods for modulating the
 CC interaction of BLR1 with its ligand, B lymphocyte chemotactant
 CC (BLC, see AAY06642). The methods comprise combining BLR1 and BLC
 CC polypeptides with a candidate modulator agent under conditions
 CC whereby, but for the presence of the agent, the polypeptides engage
 CC in a first interaction, and determining a second interaction of the
 CC polypeptides in the presence of the agent, wherein a difference

CC between the first and second interactions indicates that the agent
CC modulates the interaction of the polypeptides. The modulator is
CC preferably an antagonist, especially dominant negative, form of
CC BLC. BLC and BLC agonists and antagonists may be useful for
CC treating viral (e.g. HIV) infections, lymphoma, B lineage
CC leukemia, and autoimmune diseases such as rheumatoid arthritis,
CC thyroiditis and diabetes.
XX
SO Sequence 372 AA;

Query Match: 96.7%; Score 29; DB 20; Length 372;
Best Local Similarity 80.0%; Pred. No. 4e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 1 FPXWR 5
11 11
Db 352 fpsvr 356

Search completed: February 27, 2002, 11:51:50
Job time: 627 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:52:58 : Search time 34.86 Seconds
(without alignments)
10.926 Million cell updates/sec

Title: US-09-446-109a-23
Perfect score: 30
Sequence: 1 FPXMR 5

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	96.7	73	1 N20H1	long neurotoxin 1
2	29	96.7	113	2 T07855	translation elonga
3	29	96.7	157	2 T24380	hypothetical prote
4	29	96.7	245	2 D82680	conserved hypothet
5	29	96.7	256	2 D85510	probable EC 3.5. a
6	29	96.7	259	2 A29831	heat-labile entero
7	29	96.7	273	2 T05454	hypothetical prote
8	29	96.7	292	2 G82960	hypothetical prote
9	29	96.7	327	2 S56162	MDCK15 protein - h
10	29	96.7	332	2 J70585	minor endoglucanas
11	29	96.7	349	2 G02913	sperm CD46 - human
12	29	96.7	359	2 I40696	endoglucanase - Ce
13	29	96.7	369	2 I57998	membrane cofactor
14	29	96.7	372	2 S26667	G protein-coupled
15	29	96.7	374	2 I54479	membrane cofactor
16	29	96.7	384	2 S01896	protein p25C20.9 l
17	29	96.7	417	2 D86251	aldehyde dehydroge
18	29	96.7	479	2 T35441	probable aldehyde
19	29	96.7	497	2 D83628	neurodisulfide re
20	29	96.7	555	2 F69312	hypothetical prote
21	29	96.7	625	2 H70330	iron(II) transport
22	29	96.7	642	2 G64605	hypothetical prote
23	29	96.7	650	2 T22002	hypothetical prote
24	29	96.7	735	2 A83006	hypothetical prote
25	29	96.7	1451	2 T40325	dermoecrotic toxin
26	29	96.7	68	2 S75058	transposase ssr289
27	29	96.7	84	2 S76091	hypothetical prote
28	29	96.7	84	2 S76091	hypothetical prote
29	29	96.7	84	2 S76091	hypothetical prote

30	28	93.3	93	2 S75008	transposase ss1192
31	28	93.3	103	2 T47718	hypothetical prote
32	28	93.3	114	2 S77061	transposase sl1066
33	28	93.3	143	1 HXK48	heat shock protein
34	28	93.3	143	1 HXK41	heat shock protein
35	28	93.3	147	1 A71560	hypothetical prote
36	28	93.3	149	2 T35846	probable integral
37	28	93.3	157	2 F81710	conserved hypothet
38	28	93.3	169	2 G69300	conserved hypothet
39	28	93.3	254	2 S76459	transposase sl1186
40	28	93.3	254	2 E82791	conserved hypothet
41	28	93.3	258	2 S45682	acelone-cyanhydrin
42	28	93.3	259	2 S76643	transposase slr051
43	28	93.3	261	2 S77171	transposase sl1171
44	28	93.3	261	2 S77351	transposase sl1171
45	28	93.3	261	2 S75081	transposase slr026

ALIGNMENTS

RESULT 1
N20H1
long neurotoxin 1 - king cobra
N:Alternate names: neurotoxin A
C:Species: Ophiophagus hannah (king cobra)
C:Date: 22-Jun-1981 #sequence_revision 22-Jun-1981 #text_change 23-Aug-1996
C:Accession: A01658
R:Joubert, P.J.
Biochim. Biophys. Acta 317, 85-98, 1973
A:Title: Snake venom toxins. The amino acid sequences of two toxins from Ophiophagus
A:Reference number: A90588; MUID:73231298
A:Accession: A01658
A:Molecule type: protein
A:Residues: 1-73 <100>
C:Superfamily: snake toxin
C:Keywords: neurotoxin; venom
F:3-21,14-42,27-31,46-57,58-63/Disulfide bonds: #status predicted

Query Match 96.7% Score 29: DB 1: Length 73:
Best Local Similarity 80.0% Pred. No. 33:
Matches 4: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

QY 1 FPXMR
DB 66 FPFWR

RESULT 2
T07855
translation elongation factor eEF-1 alpha chain - rape (fragment)
M:Alternate names: translation elongation factor EF-1a
C:Species: Brassica napus (rape)
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 31-Mar-2000
C:Accession: T07855
R:Saez-Vasquez, J.
submitted to the EMBL Data Library, February 1995
A:Reference number: Z16177
A:Accession: T07855
A:Status: preliminary: translated from GR/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-113 <SAB>
A:Cross-references: EMBL:U21744; NID:q914912; P10N:AA86366.1; P10:q914913
A:Experimental source: cv. Samourai; 5 days old etiolated seedlings
C:Superfamily: translation elongation factor Tu; translation elongation factor Tu h
C:Keywords: GTP binding; protein biosynthesis

Query Match 96.7% Score 29: DB 2: Length 113:
Best Local Similarity 80.0% Pred. No. 50:
Matches 4: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

OY 1 FPXMR 5
DB 103 FPXMR 107

RESULT 3

T24380
hypothetical protein T03D8.2 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24380
R:Mortimore, B.
Submitted to the EMBL Data Library, March 1997
A:Reference number: Z19884
A:Accession: T24380
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-157 <MIL>
A:Cross-references: EMBL:Z292838; PIDN:CA807406.1; GSPDB:GN00023; CESP:T03D8.2
A:Experimental source: clone T03D8
C:Genetics:
A:Gene: CESP:T03D8.2
A:Map position: 5
A:Introns: 49/3; 83/1; 125/3

Query Match 96.7%; Score 29; DB 2; Length 157;
Best Local Similarity 80.0%; Pred. No. 68;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
DB 40 FPXMR 44

RESULT 4

D82680
conserved hypothetical protein XF1438 [imported] - *Xylella fastidiosa* (strain 9a5c)

C:Species: *Xylella fastidiosa*
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: D82680
R:anonymous. The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
Nature 406, 151-157, 2000
A>Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
A:Reference number: A82515; MUID:20365717
A>Note: for a complete list of authors see reference number A59328 below
A:Accession: D82680
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-245 <SIM>
A:Cross-references: GB:AE003974; GB:AE003849; NID:g9106454; PIDN:AAF84247.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Stimpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
B:Jones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir
as-Neto, E.; Docena, C.; El-Dorfy, H.; Facincanl, A.P.; Ferreira, A.J.S.
Submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Kleger, J.E.; Kuramae, E.E.; Laig
chado, M.A.; Medeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, B
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmiter, D.A
Rodrigues, V.; Rosa, A.J.; de M.; de Rosa Jr., V.E.; de Sa, R.C.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, F.R.; da Silva Jr., W.A.; da Silveir
M.; Tsunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1438

Query Match 96.7%; Score 29; DB 2; Length 245;
Best Local Similarity 80.0%; Pred. No. 1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
DB 197 FPXMR 201

RESULT 5

D85510
probable EC 3.5. amidase-type enzyme yafV [imported] - *Escherichia coli* (strain O157:

C:Species: *Escherichia coli*
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: D85510
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
Hiller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodu
Nature 409, 529-533, 2001
A>Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: D85510
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-256 <STO>
A:Cross-references: GB:AE005174; NID:g12512987; PIDN:AAG54544.1; GSPDB:GN00145; UMGF:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: yafV

Query Match 96.7%; Score 29; DB 2; Length 256;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
DB 244 FPXMR 248

RESULT 6

A29831
heat-labile enterotoxin IIA chain A precursor - *Escherichia coli*

R:Alternate names: LT-IIa
C:Species: *Escherichia coli*
C>Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 16-Jul-1999
C:Accession: A29831
R:Pickert, C.L.; Weinstein, D.L.; Holmes, R.K.
J. Bacteriol. 169, 5180-5187, 1987
A>Title: Genetics of type IIA heat-labile enterotoxin of *Escherichia coli*: operon *fus*
A:Reference number: A91849; MUID:88032841
A:Accession: A29831
A:Molecule type: RNA
A:Residues: 1-259 <PIC>
A:Cross-references: GB:M17894; NID:g146671; PIDN:AAA24093.1; PID:g146672
A>Note: the authors translated the codon TAT for residue 225 as Thr
C:Superfamily: heat-labile enterotoxin chain A
R:1-18/Domain: signal sequence #status predicted <SIG>
F:19-259/Product: heat-labile enterotoxin IIA chain A #status predicted <LTA>

Query Match 96.7%; Score 29; DB 2; Length 259;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXMR 5
DB 187 FPXMR 191

RESULT 7

T05454
hypothetical protein F7K2.180 - *Arabidopsis thaliana*

C:Species: *Arabidopsis thaliana* (mouse-ear cress)
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Sep-1999
C:Accession: T05454
R:Bevan, M.; Medler, H.; Wambutt, R.; Bancroft, I.; Meyers, R.; May, K.F.X.; Schme

submitted to the Protein Sequence Database, November 1998
A:Reference number: Z15416

A:Accession: T05454

A:Molecule type: DNA

A:Residues: 1-273 <BEV>

A:Cross-references: EMBL:AL033545

A:Experimental source: cultivar Columbia; BAC clone F7K2

C:Genetics:

A:Map position: 4

A:Note: F7K2.180

C:Superfamily: Arabidopsis hypothetical protein F7K2.180

Query Match

Best Local Similarity 96.7%; Score 29; DB 2; Length 273;
Matches 4; Conservativity 80.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FPXMR 5

Db 79 FPSMR 83

RESULT 8

C82960
hypothetical protein PA5488 [Imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: C82960

R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warener, P.; Hickey, M.J.; B.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim,
.: Lory, S.; Olson, M.V.

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337

A:Accession: C82960

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-292 <STO>

A:Cross-references: GB:AE004961; GB:AE004091; NID:g9951814; PIDN:AGC08873.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA5488

Query Match

Best Local Similarity 96.7%; Score 29; DB 2; Length 292;
Matches 4; Conservativity 80.0%; Pred. No. 1.2e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FPXMR 5

Db 231 FPSMR 235

RESULT 9

S56162

MDCR15 protein - human

C:Species: Homo sapiens (man)

C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999

C:Accession: S56162

R:Barcella, L.; Loetscher, M.; Tobler, A.; Baggiolini, M.; Moser, B.

Biochem. J. 309, 773-779, 1995
A:Title: Sequence variation of a novel heptahelical leucocyte receptor through alternat
A:Reference number: S56162; MUID:9536951

A:Accession: S56162

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-327 <BAR>

A:Cross-references: EMBL:X68829; NID:g840783; PIDN:CAA48723.1; PID:g840784

C:Superfamily: vertebrate rhodopsin

Query Match

Best Local Similarity 96.7%; Score 29; DB 2; Length 327;
Matches 4; Conservativity 80.0%; Pred. No. 1.4e+02;

Oy 1 FPXMR 5

Db 231 FPSMR 235

RESULT 10

JT0585

minor endoglucanase (EC 3.2.1.-) precursor - Erwinia chrysanthemi

C:Species: Erwinia chrysanthemi

C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 08-Oct-1999

C:Accession: JT0585; JT0586

R:Guiseppi, A.; Aymeric, J.L.; Caml, B.; Barras, F.; Creuzel, N.

Gene 106, 109-114, 1991

A:Title: Sequence analysis of the cellulase-encoding cely gene of Erwinia chrysanthem

A:Reference number: JT0585; MUID:92039050

A:Accession: JT0585

A:Molecule type: DNA

A:Residues: 1-332 <GUT>

A:Cross-references: GB:M74044; NID:g148391; PIDN:AAA24818.1; PID:g148392

A:Experimental source: strain 3937

A:Accession: JT0586

A:Molecule type: protein

A:Residues: 24-33 <GUT1>

C:Genetics:

A:Gene: cely

C:Keywords: glycosidase; hydrolase

F:1-23/Domain: signal sequence #status predicted <SIG>

F:24-332/Product: minor endoglucanase #status predicted <MIN>

Query Match

Best Local Similarity 96.7%; Score 29; DB 2; Length 332;
Matches 4; Conservativity 80.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FPXMR 5

Db 178 FPSMR 182

RESULT 11

G02913

Sperm CD46 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 09-Jun-2000

C:Accession: G02913

R:Hara, T.

submitted to GenBank, March 1996

A:Reference number: H01942

A:Accession: G02913

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-349 <HAR>

A:Cross-references: GB:D84105; NID:g1256700; PIDN:BA12224.1; PID:g1256701

C:Superfamily: herpesvirus complement control protein; complement factor H repeat hom

F:35-94/Domain: complement factor H repeat homology <FH01>

F:99-157/Domain: complement factor H repeat homology <FH02>

F:162-223/Domain: complement factor H repeat homology <FH03>

F:228-283/Domain: complement factor H repeat homology <FH04>

Query Match

Best Local Similarity 96.7%; Score 29; DB 2; Length 349;
Matches 4; Conservativity 80.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FPXMR 5

Db 11 FPSMR 15

RESULT 12

I40696

endoglucanase - Cellulomonas uda
 C:Species: Cellulomonas uda
 C:Date: 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 05-Nov-1999
 C:Accession: 140696
 R:Nakamura, K.; Mitsuwa, N.; Kitamura, K.
 J. Biotechnol. 4, 247-254, 1986
 A:Title: Sequence of a cellulase gene of Cellulomonas uda CB4.
 A:Reference number: 140696
 A:Accession: 140696
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-359 <RES>
 A:Cross-references: GB:M36503; NID:g144430; PIDN:AAA23090.1; PID:g144431

Query Match 96.7%; Score 29; DB 2; Length 359;
 Best Local Similarity 80.0%; Pred. No. 1.5e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPMXR 5
 DB 178 FPMXR 182

RESULT 13
 157998
 membrane cofactor protein - human
 C:Species: Homo sapiens (man)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 19-May-2000
 C:Accession: 157998
 R:Cervoni, F.; Fentchel, P.; Akhond, C.; Hsi, B.L.; Rossi, B.
 Mol. Reprod. Dev. 34, 107-113, 1993
 A:Title: Characterization of a cDNA clone coding for human testis membrane cofactor protein
 A:Reference number: 157998; MUID:93119658
 A:Accession: 157998
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-369 <RES>
 A:Cross-references: GB:S51940; NID:g262937; PIDN:AB24802.1; PID:g262938
 C:Superfamily: herpesvirus complement control protein; complement factor H repeat homolog
 F:35-94/Domain: complement factor H repeat homology <FH01>
 F:99-157/Domain: complement factor H repeat homology <FH02>
 F:162-223/Domain: complement factor H repeat homology <FH03>
 F:228-283/Domain: complement factor H repeat homology <FH04>

Query Match 96.7%; Score 29; DB 2; Length 369;
 Best Local Similarity 80.0%; Pred. No. 1.5e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPMXR 5
 DB 11 FPMXR 15

RESULT 14
 S26667
 G protein-coupled receptor BURL - human
 C:Species: Homo sapiens (man)
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jul-2000
 C:Accession: S26667
 R:Dodner, T.; Woll, I.; Emrich, T.; Lipp, M.
 Eur. J. Immunol. 22, 2795-2799, 1992
 A:Title: Differentiation-specific expression of a novel G protein-coupled receptor from
 A:Reference number: S26667; MUID:93049615
 A:Accession: S26667
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-372 <DOB>
 A:Cross-references: EMBL:X68149; NID:g29459; PIDN:CAA48252.1; PID:g29460
 C:Genetics:
 A:Gene: GDB:HLRI
 A:Cross-references: GDB:136235; OMIM:601613

A:Map position: 15q26.1-15q26.1
 C:Superfamily: vertebrate rhodopsin
 C:Keywords: G protein-coupled receptor; transmembrane protein

Query Match 96.7%; Score 29; DB 2; Length 372;
 Best Local Similarity 80.0%; Pred. No. 1.5e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPMXR 5
 DB 352 FPMXR 356

RESULT 15
 154479
 membrane cofactor protein precursor, splice form pm5.1 - human
 N:Alternate names: lymphocyte surface glycoprotein CD46
 C:Species: Homo sapiens (man)
 C:Date: 07-Jun-1996 #sequence_revision 07-Jun-1996 #text_change 31-Mar-2000
 C:Accession: 154479
 R:Purcell, D.F.; Russell, S.M.; Deacon, N.J.; Brown, M.A.; Hooker, D.J.; McKenzie, I.
 Immunogenetics 33, 335-344, 1991
 A:Title: Alternatively spliced RNAs encode several isoforms of CD46 (MCP), a regulator
 A:Reference number: 154479; MUID:91267562
 A:Accession: 154479
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-377 <RES>
 A:Cross-references: GB:M58050; NID:g180136; PIDN:AAA62833.1; PID:g180137
 C:Genetics:
 A:Gene: GDB:MCP
 A:Cross-references: GDB:120169; OMIM:120920
 R:Map position: 1q32-1q32
 C:Function:
 A:Description: for the factor I-mediated cleavage of the complement convertases
 C:Superfamily: herpesvirus complement control protein; complement factor H repeat homolog
 F:35-94/Domain: complement factor H repeat homology <FH01>
 F:99-157/Domain: complement factor H repeat homology <FH02>
 F:162-223/Domain: complement factor H repeat homology <FH03>
 F:228-283/Domain: complement factor H repeat homology <FH04>
 F:329-351/Domain: transmembrane #status predicted <TM>

Query Match 96.7%; Score 29; DB 2; Length 377;
 Best Local Similarity 80.0%; Pred. No. 1.6e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPMXR 5
 DB 11 FPMXR 15

Search completed: February 27, 2002, 11:52:58
 Job time: 450 sec

6

DE DISULFIDE BOND FORMATION PROTEIN B (DISULFIDE OXIDOREDUCTASE)
 DE (FRAGMENT).
 GN DSB. *Vibrio alginolyticus*.
 GN Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; *Vibrio*.
 OX NCBI_TaxID=663;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-B138-2;
 RA Nakamura T., Enomoto H., Unemoto T.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: REQUIRED FOR DISULFIDE BOND FORMATION IN SOME
 CC PERIPLASMIC PROTEINS. ACTS BY OXIDIZING THE DSB PROTEIN (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
 CC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE DSB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D83728; BAA12087.1; -
 DR InterPro: IPR003752; DSB.
 DR Pfam: PF02600; DsbB.1.
 KM Oxidoreductase; Redox-active center; Electron transport; Chaperone;
 KM Transmembrane; Inner membrane.
 KW DOMAIN 3 16 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 17 33 POTENTIAL.
 FT DOMAIN 34 51 PERIPLASMIC (POTENTIAL).
 FT TRANSSEM 52 67 POTENTIAL.
 FT DOMAIN 68 74 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 75 92 POTENTIAL.
 FT DOMAIN 93 147 PERIPLASMIC (POTENTIAL).
 FT TRANSSEM 148 165 POTENTIAL.
 FT DISULFID 43 46 REDOX-ACTIVE (BY SIMILARITY).
 FT DISULFID 107 133 REDOX-ACTIVE (BY SIMILARITY).
 FT NON_TER 165 165
 SQ SEQUENCE 165 AA: 18745 MW: 43396 D5AA3272D CRC64;
 Query Match 96.7%; Score 29; DB 1; Length 165;
 Best Local Similarity 80.0%; Pred. No. 37;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FPXWR 5
 DB 113 FPMWR 117
 RESULT 3
 E2AA_ECOLI STANDARD; PRT: 259 AA.
 ID E2AA_ECOLI STANDARD; PRT: 259 AA.
 AC P13810;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE HEAT-LABILE ENTEROTOXIN IIA, A CHAIN PRECURSOR (LT-IIA).
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC *Escherichia*.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88032841; PubMed=2822667;
 RA Pickett C.L., Weinstein D.L., Holmes R.K.;
 RT "Genetics of type IIA heat-labile enterotoxin of *Escherichia coli*:
 RT operon fusions, nucleotide sequence, and hybridization studies.";
 RL J. Bacteriol. 169:5180-5187(1987).

CC -1- FUNCTION: THE BIOLOGICAL ACTIVITY OF THE TOXIN IS PRODUCED BY
 CC THE A CHAIN, WHICH ACTIVATES INTRACELLULAR ADENYL CYCLASE.
 CC -1- SUBUNIT: HETEROHOMER OF ONE A CHAIN AND OF FIVE B CHAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M17894; AAA24093.1; -
 DR PIR: A29631; A29831.
 DR HSSP: P43528; ITIL.
 DR InterPro: IPR001144; Enterotoxin_A.
 DR Pfam: PF01375; Enterotoxin_A.1.
 DR PRINTS: PR00771; ENTEROTOXIN_A.
 KM Enterotoxin; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 259 HEAT-LABILE ENTEROTOXIN IIA, A CHAIN.
 FT DISULFID 203 215 BY SIMILARITY.
 FT ACT_SITE 128 128
 SQ SEQUENCE 259 AA: 29242 MW: 9965311A32CABEAA CRC64;
 Query Match 96.7%; Score 29; DB 1; Length 259;
 Best Local Similarity 80.0%; Pred. No. 56;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FPXWR 5
 DB 187 FPMWR 191
 RESULT 4
 GUNY_ERWCH STANDARD; PRT: 332 AA.
 ID GUNY_ERWCH STANDARD; PRT: 332 AA.
 AC P27032;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE MINOR ENDOGLUCANASE Y PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE
 DE Y) (CELLULOSE Y) (EGY).
 GN CELY.
 OS *Erwinia chrysanthemi*.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC *Pectobacterium*.
 OX NCBI_TaxID=556;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-33.
 RC STRAIN=3937;
 RX MEDLINE=92039050; PubMed=1937031;
 RA Guiseppe A., Aymeric J.-L., Cami B., Barras F., Creuzet N.;
 RT "Sequence analysis of the cellulase-encoding cell gene of *Erwinia*
 RT *chrysanthemi*: a possible case of interspecies gene transfer.";
 RL Gene 106:109-114(1991).
 CC -1- FUNCTION: REPRESENTS ONLY 3% OF THE GLOBAL CELLULOSE ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
 CC LINKAGES IN CELLULOSE.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY D (FAMILY 8 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M74044; AAA24818.1; -
 DR PIR: JT0585; JT0585.

DR InterPro: IPR002037; Glyco_hydro_8.
DR Pfam: PF01270; Glyco_hydro_8; 1.
DR PRINTS: PR00735; GLYHDLASE8.
DR PROSITE: PS00812; GLYCOSYL_HYDROL_F8; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 23
FT CHAIN 24 332 MINOR ENDOGLUCANASE Y.
FT ACT_SITE 53 53 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 110 110 NUCLEOPHILE (POTENTIAL).
SQ SEQUENCE 332 AA: 37592 MW: 654760428227079 CRC64:
Query Match: 96.7%; Score 29; DB 1; Length 332;
Best Local Similarity 80.0%; Pred. No. 71;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 FPMXR 5
DB 178 FPMXR 182
RESULT 5
ID GUN_CELUD STANDARD: PRT: 359 AA.
AC P18336:
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE ENDOGLUCANASE PRECURSOR (EC 3.2.1.4) (ENDO-1,4-BETA-GLUCANASE)
DE (CELLULASE).
OS Cellulomonas uda.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococccineae; Cellulomonadaceae; Cellulomonas.
OX NCBI_TaxID=1714;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-37.
RC STRAIN=C84;
RA Nakamura K., Misawa N., Kitamura K.;
RL "Sequence of a cellulase gene of Cellulomonas uda C84.";
RL J. Biotechnol. 4:247-254(1986).
CC -1- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
CC (2) EXOCELLULOBIOSYLASES THAT CUT THE DISACCHARIDE CELLULOSE
CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER
CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GLUCOSIDIC
CC LINKAGES IN CELLULOSE.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY D (FAMILY 8 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M36503; AAA23090.1; -
DR InterPro: IPR002037; Glyco_hydro_8.
DR Pfam: PF01270; Glyco_hydro_8; 1.
DR PROSITE: PS00812; GLYCOSYL_HYDROL_F8; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 23
FT CHAIN 24 359 ENDOGLUCANASE.
FT ACT_SITE 53 53 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 110 110 NUCLEOPHILE (POTENTIAL).
SQ SEQUENCE 359 AA: 40690 MW: 044507571868148 CRC64:
Query Match 96.7%; Score 29; DB 1; Length 359;

Best Local Similarity 80.0%; Pred. No. 76;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 FPMXR 5
DB 178 FPMXR 182
RESULT 6
ID CCR5_HUMAN STANDARD: PRT: 372 AA.
AC P3302; 014811;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE C-X-C CHEMOKINE RECEPTOR TYPE 5 (CXCR-5) (CXCR-5) (BURKITT'S LYMPHOMA
DE RECEPTOR 1) (MONOCYTE-DERIVED RECEPTOR 15) (MDR15).
DE BLR1 OR CXCR5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (LONG FORM).
RC TISSUE=Lymphocytes;
RX MEDLINE=93049615; PubMed=1425907;
RA Donner T., Wolf I., Emlich T., Lipp M.;
RT "Differentiation-specific expression of a novel G protein-coupled
RT receptor from Burkitt's lymphoma.";
RL Eur. J. Immunol. 22:2795-2799(1992).
RN [2]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Blood.
RX MEDLINE=95366951; PubMed=7639692;
RA Barella L., Joelscher M., Tobler A., Baggiolini M., Moser B.;
RT "Sequence variation of a novel heptahelical leucocyte receptor
RT Blochem. J. 309:773-779(1995).
RN [3]
RP LIGAND BINDING.
RX MEDLINE=98130629; PubMed=9463416;
RA Legler D.F., Joelscher M., Stuber Roos R., Clark-Lewis I.,
RA Baggiolini M., Moser B.;
RT "B cell-attracting chemokine 1, a human CXCR chemokine expressed in
RT lymphoid tissues, selectively attracts B lymphocytes via
RT BLR1/CXCR5.";
RL J. Exp. Med. 187:655-660(1998).
CC -1- FUNCTION: CYTOKINE RECEPTOR THAT BINDS TO BLR1 EXERTS
CC POSSIBLY A REGULATORY FUNCTION IN BURKITT LYMPHOMA (BL)
CC LYMPHOMAGENESIS AND/OR B-CELL DIFFERENTIATION. IT IS A POTENTIAL
CC CANDIDATE FOR CELL-CELL INTERACTION, AND ACTIVATION OF MATURE B-
CC LYMPHOCYTES IN LYMPHATIC TISSUES.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM: ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSION IN MATURE B-CELLS AND BURKITT
CC LYMPHOMA CELLS.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X68149; CA48252.1; -
DR EMBL: X68829; CA48723.1; -
DR PIR: S26667; S26667.
DR GCRDB: GCR_0453; -
DR GCRDB: GCR_2072; -
DR GCRDB: GCR_2612; -

DR MIM: 601613; -
 DR InterPro: IPR000276; GPCR_Rhodpsn.
 DR Pfam: PF00001; 7cm_1; 1.
 DR PRINTS: PR00237; GPCRHHODPSN.
 DR PRINTS: PR00564; BURKITSLYM.
 DR PROSITE: PS00237; G_PROTEIN_RECIP_FL_1; 1.
 DR PROSITE: PS50262; G_PROTEIN_RECIP_FL_2; 1.
 DR G-protein coupled receptor: Transmembrane; Glycoprotein; B-cell.
 KW Alternative splicing.
 FT DOMAIN 1 55 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 56- 76 1 (POTENTIAL).
 FT DOMAIN 77 88 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 89 109 2 (POTENTIAL).
 FT DOMAIN 110 124 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 125 145 3 (POTENTIAL).
 FT DOMAIN 146 167 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 168 188 4 (POTENTIAL).
 FT DOMAIN 189 219 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 220 240 5 (POTENTIAL).
 FT DOMAIN 241 259 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 260 280 6 (POTENTIAL).
 FT DOMAIN 281 304 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 305 325 7 (POTENTIAL).
 FT DOMAIN 326 372 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 28 28 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DISULFID 122 202 POTENTIAL.
 FT VARSPPLIC 1 45 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 344 344 G -> S (IN REF. 2).
 SO SEQUENCE 372 AA: 41955 MW: 60840839492ACCF CRC64:

Query Match 96.7%; Score 29; DB 1; Length 372;
 Best Local Similarity 80.0%; Pred. No. 79;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PPMR 5
 11 11
 DB 352 PPMR 356

RESULT 7
 MCP_HUMAN
 ID MCP_HUMAN STANDARD; PRT: 377 AA.
 AC P15529;
 DT 01-APR-1990 (rel. 14, Created)
 DT 01-OCT-1996 (rel. 34, Last sequence update)
 DT 20-AUG-2001 (rel. 40, Last annotation update)
 DE MEMBRANE COFACTOR PROTEIN PRECURSOR (CD46 ANTIGEN) (TROPHOBLAST
 DE LEUCOCYTE COMMON ANTIGEN) (TLX).
 GN MCP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxId:9606;
 RN 11
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 35-58.
 RX MEDLINE-88286080; PubMed-3260937;
 RA Lublin D.M., Liszewski M.K., Post T.W., Arce M.A., Le Beau M.M.,
 RA Reberich M.B., Lemons R.S., Seta T., Atkinson J.P.;
 RT "Molecular cloning and chromosomal localization of human membrane
 RT cofactor protein (MCP). Evidence for inclusion in the multigene
 RT family of complement-regulatory proteins.";
 RL J. Exp. Med. 168:181-194(1988).
 [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-TESTIS;
 RX MEDLINE-93119658; PubMed-8418811;
 RA Ceroni F., Penicel P., Akhoundi C., Hsi B.L., Rossi B.;
 RT "Characterization of a cDNA clone coding for human testis membrane
 RT cofactor protein (MCP, CD46).";
 RL Mol. Reprod. Dev. 34:107-113(1993).
 [3]

RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RX MEDLINE-91267562; PubMed-2050389;
 RA Purcell D.F., Russell S.M., Deacon N.J., Brown M.A., Hooker D.J.,
 RA McKenzie I.F.;
 RT "Alternatively spliced RNAs encode several isoforms of CD46 (MCP), a
 RT regulator of complement activation.";
 RL Immunogenetics 33:335-344(1991).
 [4]
 RP SEQUENCE OF 1-34 FROM N.A.
 RX MEDLINE-94014356; PubMed-7691939;
 RA Cui W., Hourcade D., Post T., Greenlund A.C., Atkinson J.P.,
 RA Kumar V.;
 RT "Characterization of the promoter region of the membrane cofactor
 RT protein (CD46) gene of the human complement system and comparison to
 RL J. Immunol. 151:4137-4146(1993).
 [5]
 RP ALTERNATIVE SPLICING.
 RX MEDLINE-92289809; PubMed-1601037;
 RA Russell S.M., Sparrow R.L., McKenzie I.F.C., Purcell D.F.J.;
 RT "Tissue-specific and allelic expression of the complement regulator
 RT CD46 is controlled by alternative splicing.";
 RL Eur. J. Immunol. 22:1513-1518(1992).
 CC - FUNCTION: MCP IS AN IMPORTANT MEMBRANE PROTEIN FOR PROTECTING HOST
 CC CELLS FROM DAMAGE BY COMPLEMENT. MCP POSSESSES COFACTOR ACTIVITY
 CC FOR INACTIVATION OF C3B AND C4B BY SERUM FACTOR I. THIS WOULD
 CC ALLOW MCP TO REGULATE THE COMPLEMENT SYSTEM AT THE LEVEL OF THE C3
 CC CONVERTASES. MAY BE INVOLVED IN THE FUSION OF THE SPERMATOZOA WITH
 CC THE OOCYTE DURING FERTILIZATION. MAY ACT BY PREVENTING COMPLEMENT
 CC ACTIVATION AT THE FETO-MATERNAL INTERFACE ON THE
 CC SYNCTIOTROPHOBLAST LAYER OF PLACENTA.
 CC - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC - ALTERNATIVE PRODUCTS: AT LEAST 14 FORMS ARE PRODUCED BY
 CC ALTERNATIVE SPLICING. THE FORM SHOWN IS THE TRANSCRIPT A. THE
 CC GAMMA ISOFORM (TRANSCRIPTS A AND B) IS PREFERENTIALLY EXPRESSED IN
 CC EBV-B CELLS AND LEUKEMIC CELLS. THE 66 KDA ALPHA ISOFORM AND THE
 CC 56 KDA BETA ISOFORM FOUND IN ALL TISSUES EXCEPT SPERM CORRESPOND
 CC TO THE TRANSCRIPTS C AND D, AND THE TRANSCRIPTS E AND F
 CC RESPECTIVELY. THE EXON 9 IS SPECIFICALLY DELETED IN SOME
 CC PLACENTAE. THE EPSILON ISOFORM CORRESPONDS TO THE TRANSCRIPTS I
 CC AND J. SPERMATOZOA DELTA ISOFORM (TRANSCRIPT N) DELETE BOTH EXONS
 CC 12 AND 13. ALL TISSUES DIFFERENTIALLY SPLICE EXON 13.
 CC - TISSUE SPECIFICITY: EXPRESSED BY ALMOST ALL CELLS EXCEPT
 CC ERYTHROCYTES AND SOME BONE MARROW CELLS.
 CC - PTM: EXTENSIVELY O-GLYCOSYLATED IN THE SER/THR-RICH DOMAIN
 CC (PROBABLY).
 CC - SIMILARITY: CONTAINS 4 SUSHI (SCR) DOMAINS.
 CC - DATABASE: NAME-PROW; NOTE-CD guide CD46 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd46.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: Y00651; CAA68675.1; -
 CC EMBL: S51940; AAB24802.1; -
 CC EMBL: M58050; AAB2833.1; -
 CC EMBL: A18585; CAA01400.1; -
 CC EMBL: S65879; AAD13968.1; -
 CC FIR: S01896; S01896.
 CC HSSP: P10998; IYVC.
 CC MIM: 120920; -
 CC InterPro: IPR000436; Sush1_SCR_CCP.
 CC Pfam: PF00084; Sush1; 4.
 CC SMART: SM00032; CCP; 4.
 CC Complement pathway: Glycoprotein; Transmembrane; Repeat; Signal;
 CC Sush1: Alternative splicing.
 CC SIGNAL 1 34
 CC CHAIN 35 377 MEMBRANE COFACTOR PROTEIN.

```

FT DOMAIN 35 328 EXTRACELLULAR (POTENTIAL).
FT TRAMSM 329 351 POTENTIAL.
FT DOMAIN 352 377 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 35 95 SUSH1 1.
FT DOMAIN 98 158 SUSH1 2.
FT DOMAIN 161 224 SUSH1 3.
FT DOMAIN 227 284 SUSH1 4.
FT DOMAIN 287 311 SER/THR-RICH.
FT DISULFID 35 80 BY SIMILARITY.
FT DISULFID 64 94 BY SIMILARITY.
FT DISULFID 99 141 BY SIMILARITY.
FT DISULFID 127 157 BY SIMILARITY.
FT DISULFID 162 210 BY SIMILARITY.
FT DISULFID 191 223 BY SIMILARITY.
FT DISULFID 228 270 BY SIMILARITY.
FT DISULFID 256 283 BY SIMILARITY.
FT CARBOHYD 83 83 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 114 114 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 273 273 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT VARSPLIC 286 300 MISSING (IN A SECOND FORM).
FT VARSPLIC 340 352 MISSING (IN ISOFORM M).
FT VARSPLIC 353 361 YLORRRKKK -> DIFKGRKKQKVELNMLTRLNQLQ
OSREAE (IN ISOFORM M).
FT VARSPLIC 362 377 TYLDETREREKETS -> KADGAETATYQTKSTPAEQ
RO (IN ISOFORM B/D/F/H/J/L).
FT VARSPLIC 362 377 MISSING (IN ISOFORM M AND ISOFORM N).
FT VARSPLIC 340 361 MISSING (IN ISOFORM N).
FT VARSPLIC 339 339 I -> IGRQWELNMLTRLNQLQOSREAE (IN
ISOFORM N).
SQ SEQUENCE 377 AA: 42247 MW: 2CA6F61752570B57 CRC64:

```

```

Query Match 96.7%; Score 29; DB 1; Length 377;
Best Local Similarity 80.0%; Pred. No. 80;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPMR 5
11 11
DB 11 FPMR 15

RESULT 8
2229_HUMAN STANDARD: PRT: 420 AA.
AC 09JUN7;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ZINC FINGER PROTEIN 229 (FRAGMENT).
CN ZNF229.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RA Shannon M., Branscomb E., Hauser L., Gordon L., Ashworth L.,
RA Stubbs L.;
RT Differential expansion of homologous zinc-finger gene families in
RT human chromosome 19q13.2 and mouse chromosome 7.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN 12
RP SEQUENCE FROM N.A.
RA Kodoyianni V., Ge Y., Krummel G.K., Kvikstad E., Grable L.,
RA Severin J., Gordon L., Shannon M., Brower A., Olsen A.S., Smith L.M.;
RT Sequence analysis of a 1mb region in 19q13.2 containing a zinc finger
RT gene cluster.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE KRUEPPL FAMILY OF C2H2-TYPE ZINC-
CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF192979; AAF07964.1;
DR EMBL: AC084239; AAG33970.1;
DR InterPro: IPR001909; KRAB.
DR InterPro: IPR000822; Znf-C2H2.
DR Pfam: PF01352; KRAB; 1.
DR Pfam: PF00096; Zf-C2H2; 2.
DR PRINTS: PR00048; ZINC_FINGER.
DR SMART: SM00349; KRAB; 1.
DR SMART: SM00355; ZNF_C2H2; 2.
DR PROSITE: PS50805; KRAB; 1.
DR PROSITE: PS50028; ZINC_FINGER_C2H2_1; 2.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 3.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 34 108 KRAB.
FT DOMAIN 349 108 ZINC_FINGERS.
FT ZN_FING 349 371 ZINC_FINGERS.
FT ZN_FING 377 399 C2H2-TYPE.
FT ZN_FING 405 >420 C2H2-TYPE.
FT NON_TER 420 420 C2H2-TYPE.
SQ SEQUENCE 420 AA: 48022 MW: FA4138BA44A2A14 CRC64:

```

```

Query Match 96.7%; Score 29; DB 1; Length 420;
Best Local Similarity 80.0%; Pred. No. 89;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPMR 5
11 11
DB 173 FPMR 177

RESULT 9
HS16_CAEEL STANDARD: PRT: 143 AA.
AC P06581; P02514;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE HEAT SHOCK PROTEIN HSP16-41.
CN HSP16-41.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Rhabditinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN 11
RP SEQUENCE FROM N.A.
RA MEDLINE=86304344; Pubmed=3017958;
RA Jones D., Russnak R.H., Kay R.J., Candido E.P.M.;
RT "Structure, expression, and evolution of a heat shock gene locus in
RT Caenorhabditis elegans (that is flanked by repetitive elements).";
RL J. Biol. Chem. 261:12006-12015(1986).
RN 12
RP SEQUENCE OF 47-143 FROM N.A.
RA MEDLINE=83220736; Pubmed=6190129;
RA Russnak R.H., Jones D., Candido E.P.M.;
RT "Cloning and analysis of cDNA sequences coding for two 16 kilodalton
RT heat shock proteins (hsps) in Caenorhabditis elegans: homology with
RT the small hsps of Drosophila.";
RL Nucleic Acids Res. 11:3187-3205(1983).
CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M14334; AAA28070.1; ALT_SEQ.
 DR EMBL; X01577; CAA25732.1; -
 DR PIR; A38884; HHKWA1.
 DR PIR; A25199; A25199.
 DR InterPro; IPR002068; Crystallin_HSP20.
 DR Pfam; PF00041; HSP20; 1.
 DR PROSITE; PS01031; HSP20; 1.
 KW Heat shock; Multigene family.
 SO SEQUENCE 143 AA; 16252 MW; C1D0F59D2636C24 CRC64;

Query Match 93.3%; Score 28; DB 1; Length 143;
 Best Local Similarity 80.0%; Pred. No. 50;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PPXMR 5
 11 11
 DB 27 PPYMR 31

RESULT 10
 HS17_CAEEL STANDARD; PRT; 143 AA.
 AC P02513;
 DT 21-JUL-1986 (Rel. 05, Created)
 DT 13-AUG-1987 (Rel. 05, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE HEAT SHOCK PROTEIN HSP16-48.
 CN HSP16-48.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 CC Rhabditidae; Pelodierinae; Caenorhabditis.
 CX NCBI_TaxID=6239;
 [1]
 [1] SEQUENCE FROM N.A. PubMed=4033652;
 RX MEDLINE=85295957; PubMed=4033652;
 RA Russnak R.H., Candido E.P.M.;
 RT "Locus encoding a family of small heat shock genes in Caenorhabditis
 RT elegans: two genes duplicated to form a 3.8-kilobase inverted
 RT repeat";
 RL Mol. Cell. Biol. 5:1268-1278(1985).
 RN [2]
 RP SEQUENCE OF 9-143 FROM N.A.
 RX MEDLINE=83220736; PubMed=6190129;
 RA Russnak R.H., Jones D., Candido E.P.M.;
 RT "Cloning and analysis of cDNA sequences coding for two 16 kilodalton
 RT heat shock proteins (hsps) in Caenorhabditis elegans: homology with
 RT the small hsps of Drosophila";
 RL Nucleic Acids Res. 11:3187-3205(1983).
 CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; K03273; AAA28069.1; -
 DR EMBL; X01576; CAA25731.1; -
 DR EMBL; K01863; AAA28064.1; -
 DR PIR; A02916; HHKWA8.
 DR PIR; A24289; A24289.
 DK InterPro; IPR002068; Crystallin_HSP20.

DR Pfam; PF00011; HSP20; 1.
 DR PROSITE; PS01031; HSP20; 1.
 KW Heat shock; Multigene family.
 SO SEQUENCE 143 AA; 16299 MW; 0D5596DFEE5B3318 CRC64;

Query Match 93.3%; Score 28; DB 1; Length 143;
 Best Local Similarity 80.0%; Pred. No. 50;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PPXMR 5
 11 11
 DB 27 PPYMR 31

RESULT 11
 HNL_MANES STANDARD; PRT; 257 AA.
 AC P52705;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE (S)-ACETONE-CYANOHYDRIN LYASE (EC 4.1.2.39) ((S)-HYDROXYNITRILE LYASE)
 DE ((S)-HYDROXYNITRILE LYASE (OXYNITRILE LYASE).
 CN HNL.
 OS Manihot esculenta (Cassava) (Manioc).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; Rosidae;
 CC eurosids I; Malpighiales; Euphorbiaceae; Manihot.
 CX NCBI_TaxID=3983;
 [1]
 [1] SEQUENCE FROM N.A. AND SEQUENCE OF 1-36 AND 168-191.
 RP TISSUE=Cotyledon;
 RX MEDLINE=94263231; PubMed=8203915;
 RA Hughes J., Carvalho F.J.P.D.C., Hughes M.A.;
 RT "Purification, characterization, and cloning of alpha-hydroxynitrile
 RT lyase from cassava (Manihot esculenta Crantz).";
 RL Arch. Biochem. Biophys. 311:496-502(1994).
 ST Arch. Biochem. Biophys. 311:496-502(1994).
 CC -1- FUNCTION: INVOLVED IN CYANOGENESIS. THE RELEASE OF HCN FROM
 CC INJURED TISSUES. DECOMPOSES A VARIETIES OF (R) OR (S) CYANOHYDRINS
 CC INTO HCN AND THE CORRESPONDING ALDEHYDES AND KETONES. THE NATURAL
 CC SUBSTRATE OF THIS ENZYME IS (S)-ACETONE CYANOHYDRIN.
 CC -1- CATALYTIC ACTIVITY: (S)-2-HYDROXYISOBUTYRONITRILE = CYANIDE +
 CC ACETONE;
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SIMILARITY: DISTANT RELATIONSHIP WITH TYPE-B
 CC CARBOXYL ESTERASE/LIPASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z29091; CAA82334.1; -
 DR HSSP; P52704; LYAS.
 DR InterPro; IPR000073; Abhydrolase.
 DR InterPro; IPR000379; Est_Lip_thioest_actsite.
 DR Pfam; PF00561; abhydrolase; 1.
 KW Lyase.
 FT ACT_SITE 79 79 BY SIMILARITY.
 FT ACT_SITE 207 207 BY SIMILARITY.
 FT ACT_SITE 235 235 BY SIMILARITY.
 SO SEQUENCE 257 AA; 29240 MW; 98B3E160ACB338C5 CRC64;

Query Match 93.3%; Score 28; DB 1; Length 257;
 Best Local Similarity 80.0%; Pred. No. 87;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5
 DB 124 FPDWR 128

RESULT 12

PABC_VIBHA STANDARD: PRT: 271 AA.

AD 036693:

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE 20-AUG-2001 (Rel. 40, Last annotation update)

DE 4-AMINO-4-DEOXYCHORISMATE LYASE (EC 4.1.3.1) (ADC LYASE) (ADCL).

GN PABC.

OS Vibrio harveyi.

OC Bacteria; Proteobacteria; gamma subdivision: Vibrionaceae; Vibrio.

OX NCBI_TaxID=669;

RN 11

RP SEQUENCE FROM N.A.

RC STRAIN=B392;

RA MEDLINE=96134997; PubMed=8550484;

RA Shen Z., Byers D.M.;

RT "Isolation of Vibrio harveyi acyl carrier protein and the fabG, acpP, and fabF genes involved in fatty acid biosynthesis."

KL J. Bacteriol. 178:571-573(1996).

CC -1- FUNCTION: CONVERTS 4-AMINO-4-DEOXYCHORISMATE INTO 4-AMINO-BENZYLATE (PABA) AND PYRUVATE.

CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).

CC -1- PATHWAY: SECOND STEP IN FOLATE BIOSYNTHESIS PATHWAY.

CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO CLASS-IV OF PYRIDOXAL-PHOSPHATE-DEPENDENT AMINOTRANSFERASES.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: U39441; AAC3592.1; -

DR InterPro: IPR001544; AminoTran_4.

DR Pfam: PF01063; AminoTran_4; 1.

DR ProDom: PD001961; AminoTran_4; 1.

DR PROSITE: PS00770; AA_TRANSFER_CLASS_4; 1.

KW Lyase: Pyridoxal phosphate: Folate biosynthesis.

FT BINDING 140 140 PYRIDOXAL PHOSPHATE (BY SIMILARITY).

SQ SEQUENCE 271 AA: 29958 MW: 3A5B5B9F398D1CD CRC64:

Query Match 93.3% Score 28; DB 1; Length 271;

Best Local Similarity 80.0%; Pred. No. 91;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5

DB 57 FPDWR 61

RESULT 13

PPNK_BACST STANDARD: PRT: 271 AA.

AC P58055:

DT 20-AUG-2001 (Rel. 40, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

DE 20-AUG-2001 (Rel. 40, Last annotation update)

DE PROBABLE INORGANIC POLYPHOSPHATE/ATP-NAD KINASE (EC 2.7.1.23)

DE (POLY(P)/ATP NAD KINASE).

GN PPNK OR NADK.

OS Bacillus stearothermophilus.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Geobacillus.

OX NCBI_TaxID=1422;

KN 11

RP SEQUENCE FROM N.A.

KA Onshima T., Sakurada H.;

KT "Inorganic polyphosphate/ATP-NAD kinase of Bacillus

KT stearothermophilus."

HL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: CATALYZES THE PHOSPHORYLATION OF NAD TO NADP. UTILIZES

CC ATP AND OTHER NUCLEOSIDE TRIPHOSPHATES AS WELL AS INORGANIC

CC POLYPHOSPHATE AS A SOURCE OF PHOSPHORUS (BY SIMILARITY).

CC -1- CATALYTIC ACTIVITY: ATP + NAD(+) = ADP + NADP(+).

CC -1- COFACTOR: REQUIRES DIVALENT METAL IONS FOR ACTIVITY (BY

CC SIMILARITY).

CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE NAD KINASE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: AB05961; BAB32727.1; -

KW Transferase: Kinase; NAD; NADP.

SQ SEQUENCE 271 AA: 30517 MW: 9E0FB427BC76662A CRC64:

Query Match 93.3% Score 28; DB 1; Length 271;

Best Local Similarity 80.0%; Pred. No. 91;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXMR 5

DB 258 FPDWR 262

RESULT 14

PAL_KLEPN STANDARD: PRT: 286 AA.

AC P37446:

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE PHOSPHOLIPASE A1 PRECURSOR (EC 3.1.1.32) (DETERGENT-RESISTANT

DE PHOSPHOLIPASE A) (DR-PHOSPHOLIPASE A) (PHOSPHATIDYLCHOLINE 1-

DE ACYLHYDROLASE) (OUTER MEMBRANE PHOSPHOLIPASE A) (OM PLA).

GN PLDA.

OS Klebsiella pneumoniae.

OC Bacteria; Proteobacteria; gamma subdivision: Enterobacteriaceae;

OC Klebsiella.

OX NCBI_TaxID=573;

RN 11

RP SEQUENCE FROM N.A.

RA MEDLINE=94131966; PubMed=8300539;

RA Brock R.G., P.M., Brinkman E., van Boxtel R., Bekkers A.C.A.P.,

RA Verheij H.M., Tommassen J.;

RT "Molecular characterization of enterobacterial plda genes encoding

RT outer membrane phospholipase A.;"

RL J. Bacteriol. 176:861-870(1994).

CC -1- FUNCTION: HYDROLYSIS OF PHOSPHATIDYLCHOLINE WITH PHOSPHOLIPASE

CC A2 (EC 3.1.1.4) AND PHOSPHOLIPASE A1 (EC 3.1.1.32) ACTIVITIES.

CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 1-ACYLGLYCERO-

CC PHOSPHOCHOLINE + A FATTY ACID ANION.

CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 2-ACYLGLYCERO-

CC PHOSPHOCHOLINE + A FATTY ACID ANION.

CC -1- COFACTOR: REQUIRES CALCIUM IONS FOR ACTIVITY.

CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE; ONE OF THE VERY FEW ENZYMES

CC LOCATED THERE.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X76901: CAA54223.1: -
 DR PIR: B36971: B36971.
 DR PIR: S40129: S40129.
 DR InterPro: IPR003187: PLA1.
 DR Pfam: PF02253: PLA1: 1.
 KM Hydrolyase: Lipid degradation: Outer membrane: Signal: Calcium.
 FT SIGNAL: 1 20 BY SIMILARITY.
 FT CHAIN: 21 286 PHOSPHOLIPASE A1.
 FT ACT_SITE: 161 161 BY SIMILARITY.
 SQ SEQUENCE 286 AA: 32544 MW: 3639F863085108A3 CRC64;

Query Match 63.3% Score 28; DB 1; Length 286;
 Best Local Similarity 80.0% Pred. NO. 96;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PPMXR 5
 Db 92 PPMXR 96

RESULT 15
 PAI_ECOLI STANDARD: PRT: 289 AA.
 AC P00631:
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PHOSPHOLIPASE A1 PRECURSOR (EC 3.1.1.32) (DETERGENT-RESISTANT
 DE PHOSPHOLIPASE A) (OR-PHOSPHOLIPASE A) (PHOSPHATIDYLCHOLINE 1-
 DE ACYLHYDROLASE) (OUTER MEMBRANE PHOSPHOLIPASE A) (OM PLA).
 CN PLDA OR B3821 OR Z5342 OR ECS4751.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria: Proteobacteria: gamma subdivision: Enterobacteriaceae:
 OC Escherichia.
 OX NCBI_TaxID=562, 83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85157492; PubMed=6397464;
 RA Homme H., Kobayashi T., Chiba N., Karasawa K., Mizushima H., Kudo I.,
 RA Inoue K., Ikeda H., Sekiguchi M., Nojima S.:
 RT "The DNA sequence encoding plid gene, the structural gene for
 RT detergent-resistant phospholipase A of E. coli.";
 RL J. Biochem. 96:1655-1664(1984).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655:
 RX MEDLINE=92358234; PubMed=1379743;
 RA Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.:
 RT "Analysis of the Escherichia coli genome: DNA sequence of the region
 RT from 84.5 to 86.5 minutes.";
 RL Science 257:771-778(1992).
 RN [3]
 RP REVISTON TO 14-15.
 RC STRAIN-K12 / MG1655:
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.:
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postal G., Hackett J., Klink S., Boulton A., Shao Y., Miller L.,
 RA Goebel E.J., Davis N.W., Lim A., Dimalanta E.T., Polomousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.:
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533(2001).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Lida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shigaawa H.:
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 RN [6]
 RP SEQUENCE OF 30-289 FROM N.A.
 RC STRAIN-K12:
 RX MEDLINE=85003590; PubMed=6383820;
 RA de Gues P., Verheij H.M., Riegan N.H., Hoekstra W.P.M., de Haas G.H.:
 RT "The pro- and mature forms of the E. coli K-12 outer membrane
 RT phospholipase A are identical.";
 RL EMBO J. 3:1799-1802(1984).
 RN [7]
 RP SEQUENCE OF 174-289 FROM N.A.
 RC STRAIN-K12:
 RX MEDLINE=87115164; PubMed=1027506;
 RA Iitno N., Nakayama K., Nakayama H.:
 RT "The recQ gene of Escherichia coli K12: primary structure and
 RT evidence for SOS regulation.";
 RL Mol. Gen. Genet. 205:298-304(1986).
 RN [8]
 RP MUTAGENESIS OF SER-172.
 RX MEDLINE=94131966; PubMed=8300539;
 RA Brock R.G.P.M., Brinkman E., van Boxtel R., Bekkers A.C.A.P.,
 RA Verheij H.M., Tommassen J.:
 RT "Molecular characterization of enterobacterial plid genes encoding
 RT outer membrane phospholipase A.";
 RL J. Bacteriol. 176:861-870(1994).
 RN [9]
 RP ACTIVE SITE SER-164, AND PARTIAL SEQUENCE.
 RX MEDLINE=91249806; PubMed=2040286;
 RA Horrevorts A.J.G., Verheij H.M., de Haas G.H.:
 RT "Inactivation of Escherichia coli outer-membrane phospholipase A by
 RT the affinity label hexadecanesulfonyl fluoride. Evidence for an
 RT active-site serine.";
 RL Eur. J. Biochem. 198:247-253(1991).
 CC -1- FUNCTION: HYDROLYSIS OF PHOSPHATIDYLCHOLINE WITH PHOSPHOLIPASE
 CC A2 (EC 3.1.1.4) AND PHOSPHOLIPASE A1 (EC 3.1.1.32) ACTIVITIES.
 CC REQUIRED FOR EFFICIENT SECRETION OF BACTERIOCIDINS. SEEMS TO BE
 CC DORMANT IN NORMAL GROWING CELLS.
 CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 1-ACYLGLYCERO-
 CC PHOSPHOCHOLINE + A FATTY ACID ANION.
 CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 2-ACYLGLYCERO-
 CC PHOSPHOCHOLINE + A FATTY ACID ANION.
 CC -1- COFACTOR: REQUIRES CALCIUM IONS FOR ACTIVITY.
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE; ONE OF THE VERY FEW ENZYMES
 CC LOCATED THERE.
 CC -1- INDUCTION: BY MEMBRANE DAMAGING, FOR EXAMPLE, BY PHAGE-INDUCED
 CC LYSIS OR TEMPERATURE SHOCK.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X02143; CAA26081.1; -.
DR EMBL: M87049; AAA67617.1; -.
DR EMBL: AE000458; AAC76824.1; -.
DR EMBL: AE005613; AAC59017.1; -.
DR EMBL: AP002567; BAB38174.1; -.
DR EMBL: M30198; AAA24516.1; -.
DR PIR: A00771; PSECA.
DR PIR: A22133; PSECA1.
DR PIR: S30711; S30711.
DR EcoGene: EG10738; plda.
DR InterPro: IPR003187; PLA1.
DR Pfam: Pf02253; PLA1; 1.
KW Hydrolase; lipid degradation; Outer membrane; Signal; Calcium;
KW Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 289 PHOSPHOLIPASE A1.
FT ACT_SITE 164 164
FT MUTAGEN 172 172 S->F: INACTIVE PROTEIN.
FT CONFLICT 14 15 LP -> FA (IN REF. 2).
FT CONFLICT 30 33 DAPA -> MTRO (IN REF. 6).
SQ SEQUENCE 289 AA: 33163 MW: A688AD32AA60F218 CRC64: .

Query Match 93.38; Score 28; DB 1; Length 289;
Best Local Similarity 80.08; Pred. No. 97;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 1 FPXWR 5
DB 95 FPLWR 99

Search completed: February 27, 2002, 11:52:17
Job time: 564 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:50:37 ; Search time 32.48 Seconds
(without alignments)
3.464 Million cell updates/sec

Title: US-09-446-109A-23

Perfect score: 30
Sequence: 1 FPXMR 5

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, AA: *
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep: *
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/PCRTUS_COMB.pep: *
6: /cgn2_6/ptodata/2/1aa/backfilest.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	96.7	38	1	US-08-176-500-118
2	29	96.7	38	1	US-08-471-052A-118
3	29	96.7	38	1	US-08-189-331-118
4	29	96.7	38	2	US-08-471-939-118
5	29	96.7	38	2	US-08-471-800-118
6	29	96.7	38	2	US-08-471-068-118
7	29	96.7	241	4	US-08-823-120-1
8	29	96.7	324	2	US-08-528-057-46
9	29	96.7	370	2	US-08-528-057-42
10	29	96.7	372	1	US-08-202-056-5
11	29	96.7	372	1	US-08-076-093A-6
12	29	96.7	372	1	US-08-701-265-6
13	29	96.7	372	1	US-08-284-586-6
14	29	96.7	372	2	US-08-805-478-6
15	29	96.7	372	2	US-08-802-627A-6
16	29	96.7	372	2	US-08-801-238-6
17	29	96.7	372	3	US-08-801-228-6
18	29	96.7	372	3	US-09-104-226-6
19	29	96.7	372	3	US-08-982-493-8
20	29	96.7	372	2	US-08-528-057-44
21	29	96.7	377	2	US-08-528-057-2
22	29	96.7	384	4	US-08-139-195-2
23	29	96.7	384	4	US-08-139-195-2
24	28	93.3	57	1	US-08-370-225-29
25	28	93.3	57	1	US-08-370-225-30
26	28	93.3	57	1	US-08-461-859-29
27	28	93.3	57	1	US-08-461-859-30

28	28	93.3	57	5	PCT-US91-10069-29	Sequence 29, Appl
29	28	93.3	57	5	PCT-US91-10069-30	Sequence 30, Appl
30	28	93.3	298	2	US-08-061-636-3	Sequence 3, Appl1
31	28	93.3	298	2	US-08-874-347-19	Sequence 19, Appl
32	28	93.3	298	3	US-09-093-522-19	Sequence 19, Appl
33	28	93.3	298	5	PCT-US94-05268-3	Sequence 3, Appl1
34	28	93.3	373	2	US-08-846-762-13	Sequence 13, Appl
35	28	93.3	374	3	US-08-962-493-6	Sequence 6, Appl1
36	28	93.3	437	2	US-09-031-059-1	Sequence 1, Appl1
37	28	93.3	437	2	US-09-031-059-3	Sequence 3, Appl1
38	28	93.3	760	1	US-08-195-152-2	Sequence 2, Appl1
39	26	86.7	10	4	US-09-461-697-409	Sequence 409, App
40	26	86.7	112	4	US-09-461-697-407	Sequence 107, App
41	26	86.7	113	1	US-07-668-648-10	Sequence 10, Appl
42	26	86.7	113	2	US-08-429-998-10	Sequence 10, Appl
43	26	86.7	113	2	US-08-431-333-10	Sequence 10, Appl
44	26	86.7	113	5	PCT-US91-02321-10	Sequence 10, Appl
45	26	86.7	862	1	US-08-325-267A-4	Sequence 4, Appl1

ALIGNMENTS

RESULT 1
US-08-176-500-118
Sequence 118, Application US/08176500
Patent No. 5498538
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 141
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennile & Edmunds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/176,500
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/013,416
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-143
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-176-500-118

Query Match 96.7%; Score 29; DB 1; Length 38;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPMXR 5
11 11
DB 10 FPMWR 14

RESULT 2

US-08-471-052A-118
Sequence 118, Application US/08471052A
Patent No. 5625033

GENERAL INFORMATION:

APPLICANT: Kay, B. K.

APPLICANT: Fowlkes, D. M.

TITLE OF INVENTION: Totally Synthetic Affinity Reagents

NUMBER OF SEQUENCES: 166

CORRESPONDENCE ADDRESS:

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/471.052A

FILING DATE: 06-JUNE-1995

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 1101-179

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212 790-9090

TELEFAX: 212 869-8864/9741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 118:

SEQUENCE CHARACTERISTICS:

LENGTH: 38 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-471-052A-118

Query Match

Best Local Similarity 96.7%; Score 29; DB 1; Length 38;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FPMXR 5
11 11
DB 10 FPMWR 14

RESULT 3

US-08-189-331-118
Sequence 118, Application US/08189331
Patent No. 5747334

GENERAL INFORMATION:

APPLICANT: Kay, B. K.

APPLICANT: Fowlkes, D. M.

TITLE OF INVENTION: Totally Synthetic Affinity Reagents

NUMBER OF SEQUENCES: 186

CORRESPONDENCE ADDRESS:

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/189,331

FILING DATE: Concurrently herewith

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 1101-155

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212 790-9090

TELEFAX: 212 869-8864/9741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 118:

SEQUENCE CHARACTERISTICS:

LENGTH: 38 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-189-331-118

OY 1 FPMXR 5
11 11
DB 10 FPMWR 14

Query Match

Best Local Similarity 96.7%; Score 29; DB 1; Length 38;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 FPMXR 5
11 11
DB 10 FPMWR 14

RESULT 4

US-08-471-939-118
Sequence 118, Application US/08471939
Patent No. 5844076

GENERAL INFORMATION:

APPLICANT: Kay, B. K.

APPLICANT: Fowlkes, D. M.

TITLE OF INVENTION: Totally Synthetic Affinity Reagents

NUMBER OF SEQUENCES: 141

CORRESPONDENCE ADDRESS:

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/471,939

FILING DATE: 06-JUN-1995

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/013,416

FILING DATE: 01-FEB-1993

ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 1101-143

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212 790-9090

TELEFAX: 212 869-8864/9741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 118:

SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-471-939-118

Query Match 96.7%; Score 29; DB 2; Length 38;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPMXR 5
1111
DB 10 FPSMR 14

RESULT 5
US-08-471-800-118
Sequence 118, Application US/08471800
Patent No. 5852167
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 141
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,800
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/013,416
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-143
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-471-800-118

Query Match 96.7%; Score 29; DB 2; Length 38;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPMXR 5
1111
DB 10 FPSMR 14

RESULT 6
US-08-471-068-118
Sequence 118, Application US/08471068
Patent No. 5948635
GENERAL INFORMATION:
APPLICANT: Kay, B. K.
TITLE OF INVENTION: Totally Synthetic Affinity Reagents
NUMBER OF SEQUENCES: 186
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,068
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/189,331
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-155
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 118:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-471-068-118

Query Match 96.7%; Score 29; DB 2; Length 38;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPMXR 5
1111
DB 10 FPSMR 14

RESULT 7
US-08-823-120-1
Sequence 1, Application US/08823120
Patent No. 6149919
GENERAL INFORMATION:
APPLICANT: Domenighini, Mario
APPLICANT: Rappuoli, Rino
APPLICANT: Pizza, Mariagrazia
TITLE OF INVENTION: Immunogenic Detoxified Mutants of
TITLE OF INVENTION: Cholera Toxin and of the Toxin Lt, Their Preparation and
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: California
COUNTRY: USA

ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/823,120
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/256,003
FILING DATE: 11-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0315,001
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 241 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-823-120-1

Query Match 96.7%; Score 29; DB 4; Length 241;
Best Local Similarity 80.0%; Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PPSMR 5
1111
DB 169 FPMR 173

RESULT 8
US-08-528-057-46
Sequence 46, Application US/08528057
Patent No. 5846715
GENERAL INFORMATION:
APPLICANT: PURCELL, Damian F. J.
APPLICANT: RUSSELL, Sarah M.
APPLICANT: MCKENZIE, Ian F. C.
TITLE OF INVENTION: CD46 VARIANTS
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/528,057
FILING DATE: CONCURRENTLY HERewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/961,686
FILING DATE: 11-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU91/00199
FILING DATE: 10-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PK0133/90
FILING DATE: 11-MAY-1990

ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 17227/112 DACO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 324 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-528-057-46

Query Match 96.7%; Score 29; DB 2; Length 324;
Best Local Similarity 80.0%; Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 PPSMR 5
1111
DB 11 FPMR 15

RESULT 9
US-08-528-057-42
Sequence 42, Application US/08528057
Patent No. 5846715
GENERAL INFORMATION:
APPLICANT: PURCELL, Damian F. J.
APPLICANT: RUSSELL, Sarah M.
APPLICANT: MCKENZIE, Ian F. C.
TITLE OF INVENTION: CD46 VARIANTS
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/528,057
FILING DATE: CONCURRENTLY HERewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/961,686
FILING DATE: 11-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU91/00199
FILING DATE: 10-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PK0133/90
FILING DATE: 11-MAY-1990
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 17227/112 DACO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 370 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-528-057-42

Query Match 96.7%; Score 29; DB 1; Length 370;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPMR 5
DB 11 FPMR 15

RESULT 10
US-08-202-056-5
Sequence 5, Application US/08202056
Patent No. 5440021
GENERAL INFORMATION:
APPLICANT: Chuntharapal, Anan
APPLICANT: Hebert, Caroline
APPLICANT: Kim, Kyung Jin
APPLICANT: Lee, James
TITLE OF INVENTION: Antibodies to Human IL-8 Type B Receptor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202.056
CLASSIFICATION: 436
FILING DATE: 25-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/677211
FILING DATE: 29-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B.
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: 706P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-202-056-5

Query Match 96.7%; Score 29; DB 1; Length 372;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPMR 5
DB 352 FPMR 356

RESULT 11
US-08-076-093A-6
Sequence 6, Application US/08076093A
Patent No. 5343503
GENERAL INFORMATION:

APPLICANT: Chuntharapal, Anan
APPLICANT: Lee, James
APPLICANT: Hebert, Caroline
APPLICANT: Jin Kim, K.
TITLE OF INVENTION: Antibodies to Human PFA Receptors
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpatin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/076.093A
CLASSIFICATION: 530
FILING DATE: 11-Jun-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/677211
FILING DATE: 29-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B.
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: 706P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: linear
US-08-076-093A-6

Query Match 96.7%; Score 29; DB 1; Length 372;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPMR 5
DB 352 FPMR 356

RESULT 12
US-08-701-265-6
Sequence 6, Application US/08701265
Patent No. 5776457
GENERAL INFORMATION:
APPLICANT: Chuntharapal, Anan
APPLICANT: Lee, James
APPLICANT: Hebert, Caroline
APPLICANT: Jin Kim, K.
TITLE OF INVENTION: Antibodies to Human PFA Receptors
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 mb floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPain (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/701.265
FILING DATE: 22-AUG-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/076093
FILING DATE: 11-Jun-1993
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/677211
FILING DATE: 29-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: 706P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-701-265-6

Query Match 96.7% Score 29; DB 1; Length 372:
Best Local Similarity 80.0% Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PFXMR 5
1111
DB 352 FPSWR 356

RESULT 13
US-08-284-586-6
Sequence 6, Application US/08284586
Patent No. 5840856
GENERAL INFORMATION:
APPLICANT: Chuncharapai, Anan
APPLICANT: Lee, James
APPLICANT: Hebert, Caroline
APPLICANT: Jin Kim, K.
TITLE OF INVENTION: Antibodies to Human PFAA Receptors
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPain (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/284.586
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/076.093A
FILING DATE: 11-Jun-1993
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/677211
FILING DATE: 29-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: 706P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-284-586-6

Query Match 96.7% Score 29; DB 2; Length 372:
Best Local Similarity 80.0% Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PFXMR 5
1111
DB 352 FPSWR 356

RESULT 14
US-08-805-478-6
Sequence 6, Application US/08805478
Patent No. 5874543
GENERAL INFORMATION:
APPLICANT: Chuncharapai, Anan
APPLICANT: Lee, James
APPLICANT: Hebert, Caroline
APPLICANT: Jin Kim, K.
TITLE OF INVENTION: ANTIBODIES TO PFAA RECEPTOR
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPain (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/805.478
FILING DATE: 25-Feb-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/284586
FILING DATE: 10-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/076093
FILING DATE: 11-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: P0706P2P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:

LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-805-478-6

Query Match 96.7% Score 29: DB 2: Length 372:
Best Local Similarity 80.0% Pred. No. 1.6e+02:
Matches 4: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

QY 1 FPXMR 5
DB 352 FPSMR 356

RESULT 15
US-08-802-627A-6
Sequence 6, Application US/08802627A
Patent No. 5892017
GENERAL INFORMATION:
APPLICANT: Lee, James
APPLICANT: Wood, William F.
TITLE OF INVENTION: NUCLEIC ACID ENCODING PFAA RECEPTOR
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Winpallin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/802,627A
FILING DATE: 19-Feb-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/284586
FILING DATE: 10-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/076093
FILING DATE: 11-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/810782
FILING DATE: 19-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Love, Richard B.
REGISTRATION NUMBER: 34,659
REFERENCE/DOCKET NUMBER: P0706P2PID2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-5530
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 372 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-802-627A-6

Query Match 96.7% Score 29: DB 2: Length 372:
Best Local Similarity 80.0% Pred. No. 1.6e+02:
Matches 4: Conservative 0: Mismatches 1: Indels 0: Gaps 0:

QY 1 FPXMR 5
DB 352 FPSMR 356

Search completed: February 27, 2002, 11:50:37
Job time: 866 sec

